# ESSAYS ON THE ROLE OF SOCIAL STATUS AND PEER EFFECTS ON INDIVIDUALS' BEHAVIOR AND WELL-BEING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Juan David Robalino

December 2017

ESSAYS ON THE ROLE OF SOCIAL STATUS AND PEER EFFECTS ON INDIVIDUALS' BEHAVIOR AND WELL-BEING

Juan David Robalino, Ph.D.

Cornell University 2017

Classic economics has long underestimated the role of social contexts on individuals' behavior. Individuals are not only driven by independent incentives but also by social influence as well as social approval and recognition. This dissertation contains three papers concerning the role of peer effects and social status on individuals' behavior and well-being.

The first chapter analyzes adolescents peer effects on cigarettes consumption by considering the popularity (i.e., social network centrality measures) of smokers and non-smokers within high schools. The analysis is based on the AddHealth dataset, which has exhaustive data on social networks in the sampled high schools and detailed information about students and household characteristics. We use variations across cohorts within schools by using school fixed effects, lagged peer's behavior at the cohort level as well as instrumental variables. We find that the popularity of smokers increases the probability of individuals smoking, while the popularity of non-smokers has the opposite effect. Analogous results apply to the number of cigarettes smoked, as well as the age of initiation. These effects persist up to fourteen years after high school.

The second chapter follows a similar strategy to analyze peer effects on high school GPA. Using AddHealth data, we find that the popularity of good students in mathematics considerably increases individuals' math GPA the following year, while the popularity of bad students has the opposite effect. The positive effect is somewhat stronger for males, yet the negative effect is much stronger for females. These patterns extend to college participation and completion. Thus, the first two chapters of this dissertation show the importance of social status in mediating the strength of peers' influence. Individuals seem to 'follow' the behavior of high-status individuals and avoid that of low-status individuals.

The final chapter looks at positional concerns in rural China with respect to rural income and remittances. More specifically, we look at how the average income and average remittances in the local county affect individuals' subjective well-being (SWB). We find that average rural income has a strong negative effect on SWB, arguably due to income comparisons. However, average remittances have a strong positive effect, possibly due to a signal effect of the potential opportunities from migration, both for potential migrants and potential remittances receivers.

## BIOGRAPHICAL SKETCH

Juan David Robalino was born in Quito, Ecuador on August $1^{st}$, 1989. He completed his MSci in Mathematics at Imperial College London, UK, in 2011. Aiming to apply mathematics to the social science, he then joined IZA –The Institute of Labor Economics– in Bonn, Germany. He started his PhD in economics at Cornell University in the summer of 2012. He is currently working at Ideas 42 in New York.

## ACKNOWLEDGEMENTS

I am extremely grateful to my committee members, Michael Macy, Robert Frank and Ori Heffetz for their advice and guidance. I am especially grateful to my advisors Michael Macy and Robert Frank who gave me unconditional support and encouragement.

I also would like to thank all the members of the Social Dynamics Lab (SDL) and the Behavioral Economics Research Group (BERG) for their constructive and valuable comments.

Finally, I am very grateful to all my family for their support throughout this journey.

## TABLE OF CONTENTS

# Peer Effects on Adolescent Smoking: Are Popular Teens more Influential?

Juan David Robalino[1,2] & and Michael Macy[3]
[1]Department of Economics, Cornell University
[2]IZA – Institute for the Study of Labor
[3]Department of Information Science, Cornell University

**Abstract**

Previous research on adolescent cigarette adoption has focused on peer influence and the perceived status gain from smoking but has ignored the status effects on peer influence. We analyze adolescent peer effects on cigarette consumption while considering the popularity of peers. The analysis is based on a four wave panel survey representative of American high school students. We measure peers' popularity by their eigenvector centrality in high school social networks. Using lagged peers' behavior, school fixed effects, and instrumental variables to control for homophily and contextual confounds, we find that the probability of smoking the following year increases with the mean popularity of smokers, while the popularity of non-smokers has the opposite effect. These effects persist seven and fourteen years later (wave 3 and 4 of the data). In addition, the probability of smoking increases with the smoking propensity of the 20% most popular teens and decreases with the smoking propensity of the bottom 80%. The results indicate the importance of knowing not only the smoking propensity within a school but also the location of smokers within the social hierarchy.

# 1 Introduction

Identifying the drivers of smoking adoption among youth remains a public health priority. Initiating smoking at a young age is correlated with smoking more cigarettes per day and with a lower probability of quitting later in life (Everett et al., 1999; Lando et al., 1999). More than 480,000 deaths are attributed to cigarette smoking every year in the US alone. At the current rate of adoption, 5.6 million Americans under 18 –about 1 of every 13– will die early from a smoking-related illness.[1]

Previous research points to peer influence as an important cause of adolescent smoking (Clark and Lohéac, 2007; Fletcher, 2010; Gaviria and Raphael, 2001; Krauth, 2006; Lundborg, 2006; Nakajima, 2007; Powell et al., 2005). Adolescents are especially vulnerable to social influence as they try to fit in with their peers (Aral and Walker, 2012; Gardner and Steinberg, 2005). Numerous studies suggest that some smokers believe that smoking promotes social status (Alexander et al., 2001; Ennett et al., 2006; Michell and Amos, 1997; Valente et al., 2005), and yet previous research has failed to test a key implication: that the propensity to smoke may increase with the popularity of smokers among peers. In other contexts, peer influence has been found to increase with their social status (Oldmeadow et al., 2003), but no previous studies have empirically tested the effects of the social status of adolescent smokers on the spread of smoking through peer networks. The need to empirically test this status-belief explanation is the starting point for our study.

# 2 Previous research on popularity and smoking

Several studies have analyzed the link between popularity of adolescents (usually measured by the number of incoming friendship nominations) and their smoking behavior. For example,

[1] Centers for Disease Control and Prevention.

Valente et al. (2005) and Ennett et al. (2006) found a positive relationship between popularity and smoking; Alexander et al. (2001) found that popular students are more likely to smoke in schools with high smoking rates, and less likely in schools with low smoking rates. Michell and Amos (1997) used qualitative data from Scottish schoolgirls and found that popular girls smoke to maintain their image while some unpopular girls smoke with the hope of gaining social status, but no effect in the mid-popularity range.

These studies indicate that smoking may be used as a strategy to climb the social ladder, based on the belief that smoking will lead to an increase in social status. However, they do not test the peer effects of smokers' social status. To address this gap, we reverse the causal arrow: instead of testing whether smoking affects status, we test how the popularity of smokers affects their influence on the behavior of their peers. We posit that the relative popularity of smokers and non-smokers will condition whether smoking will be associated with social status in that population. To the best of our knowledge, this is the first study of the effect of smokers' popularity on peer influence for smoking.[2]

It is worth noting that there have been anti-smoking field interventions that relied implicitly on the hypothesis that we are testing. For example, to promote an anti-smoking campaign in high schools, a field experiment explicitly recruited individuals nominated by fellow students as "influential" (Campbell et al., 2008), and their intervention proved successful in comparison to the smoking rates of the control high schools that did not recruit any students. In short, this research assumed status differences in influence but did not test that assumption. Our study is motivated by the need to empirically test an assumption that informs intervention strategies not only into smoking but into other public health, advertising, and electoral campaigns.

[2]In his book "The Tipping Point", Gladwell (2006) presents informal interviews from smokers describing their first impressions about cigarettes; most descriptions point towards a given person whom they admired in their youth and who smoked. In fact, this study is inspired by this book.

# 3 The reflection problem

Like most previous studies of peer influence on adolescent smoking, we rely on observational data. Seminal work by Manski (1993) identifies the inherent difficulty in estimating peer effects with observational data, which he refers to as the reflection problem. According to Manski, a correlation between the average behavior of the group and an individual's behavior can be attributed to three possible mechanisms:

- *endogenous effects*, where an individual's choice is influenced by the choices of the group, as occurs with peer effects on behavior;
- *exogenous (contextual) effects*, where individuals in a given group may behave similarly because the whole group has experienced an (unobserved) exogenous shock;
- *correlated effects*, where individuals in a group behave similarly because they have similar unobservable characteristics and self-select into the group.

The challenge with observational data is how to disentangle endogenous effects from contextual and correlated effects. Several identification approaches have been used. Sacerdote (2001) studied peer effects on the Grade Point Average (GPA) of college students using exogenous group formation from random allocation to college dorms in order to control for peers' self-selection and homophily. He also used peers' lagged GPA (from high-school) to control for contextual confounds.

In the absence of random allocation of peers, several studies of peer effects on smoking have used broad pseudo-exogenous peer groups, controls for school level fixed effects, as well as instrumental variables for peers' behavior. Lundborg (2006) used the smoking rate among classmates as peer referents; Fletcher (2010) used the school grade; Clark and Lohéac (2007) used lagged behavior at the grade level (i.e., the smoking rate during the previous year); Norton et al.

(1998), Gaviria and Raphael (2001) and Powell et al. (2005) used the whole school. To account for contextual cofounds, Gaviria and Raphael (2001), and Powell et al. (2005) controlled for cigarette prices and public policy variables, while Lundborg (2006), Clark and Lohéac (2007), and Fletcher (2010) included school fixed effects. Norton et al. (1998), Gaviria and Raphael (2001), Powell et al. (2005), and Fletcher (2010) also instrumented the behavior of peers using household or neighborhood characteristics. All these studies found strong evidence of peer influence on cigarettes use.[3]

In line with previous studies, we use the lagged behavior of peers at the school grade level to control for homophily and include school fixed effects to control for contextual confounds. Nevertheless, the observed effects from multivariate models remain susceptible to unobserved heterogeneity. We therefore also use instrumental variables to test the robustness of the causal inferences.

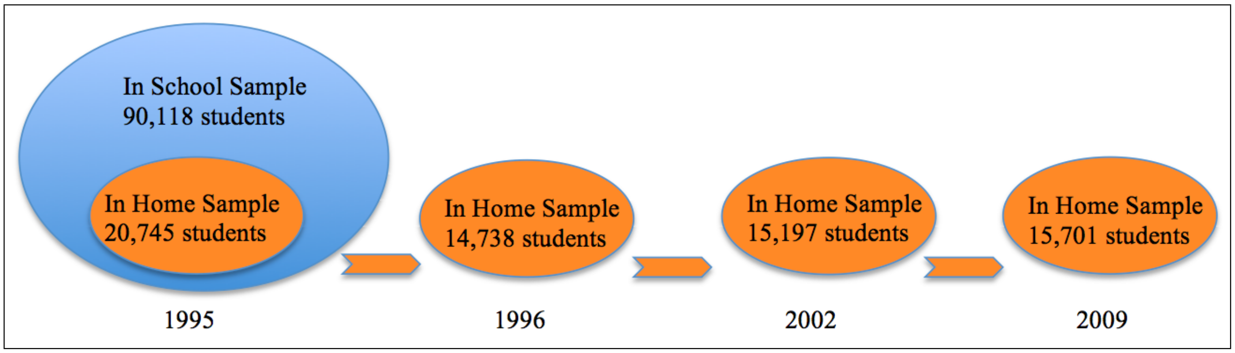
# 4 Data and methods

The present study is based on data from the National Longitudinal Survey of Adolescent Health (AddHealth), a four-wave panel (1995, 1996, 2002 and 2009) administered to a representative sample of American High Schools. This data set contains detailed information about substance use as well as rich social network data with which to measure the popularity of students. The multi-wave longitudinal data also includes identifiers for grade, year, school, and detailed mea-

[3] A more controversial approach to identify peer effects is that of Christakis and Fowler (2007): using data from the The Framingham Heart Study consisting of a large (sparse) social network, they relied on the direction of social tie nominations and the panel aspect of the data (the behavior of nominated peers at time $t$ to predict the behavior of the focal individual at time $t+1$) to capture peers' influence on the diffusion of obesity and subsequently on the diffusion of smoking (Christakis and Fowler, 2007, 2008). Although they found peer effects in both domains, their approach has been criticized for not accounting sufficiently for contextual effects and homophily (Cohen-Cole and Fletcher, 2008).

Other studies identify smoking peer effects using Monte Carlo estimations. Two main types of models exist: the first are structural models from game theory where the influence of peers is estimated as the equilibrium outcome of strategic interactions (Krauth, 2006; Nakajima, 2007); the second type consists of agent-based models capturing the co-evolution of friendship formation and peers' influence (Mercken et al., 2010; Steglich and Snijders, 2010).

Figure 1: Data structure.




sures of household and parents' characteristics. The In-School survey conducted in 1995 covered 90,118 adolescents in 144 schools. A representative sub-sample of 20,745 students also completed the 1995 In-Home wave 1 survey which included more detailed information such as family income and parents' smoking habits. Finally, about 15,000 students were surveyed again in 1996 (In-Home survey wave 2), as well as 2002 (In-Home wave 3) and 2009 (In-Home wave 4; see Fig 1). After excluding individuals with missing information and schools with insufficient social network data, we have a sample of about 66,000 peers and a core sample of about 7,500 individuals.

## 4.1 Popularity measure

AddHealth identifies peers by asking each respondent to name their five closest male friends within the school and their five closest female friends in order of importance (i.e., best friend, second best friend, etc). These nominations form the social network of the school and can then be used to measure the popularity of each respondent by their *eigenvector centrality*.[4] Intuitively, the popularity of individual $i$ will be proportional to the popularity of peers who

[4] In the working paper Robalino (2016) we obtain similar results using in-degree, out-degree, and Katz–Bonacich centrality; see Jackson (2008) Chapter 2 for a discussion about network centrality measures.

nominated $i$ as a friend, hence being nominated by popular teens makes you more popular. Formally, the centrality $v_i$ of individual $i$ in school $s$ is

$$v_i = \frac{1}{\lambda} \sum_j v_j a_{ji}$$

where $a_{ij} = 1/n_{ij}$ if individual $i$ nominated individual $j$ as his/her $n^{th}$ best friend and $a_{ij} = 0$ otherwise; or equivalently using vector notation we recover the eigenvector expression $\lambda v = A'_s v$, where $A_s$ is the so called adjacency matrix of the social network in school $s$ and $\lambda$ is the largest eigenvalue –in order to ensure that all the elements of $v$ are positive as guaranteed by the Perron–Frobenius theorem.[5] This weighted-eigenvector centrality (i.e., weighted by the inverse of the friendship nomination order) is preferred as it also takes account of the difference between being the best friend of many people and being the fifth best friend of many people.[6]

We compare the centrality measure across school/grades in two ways: by standardizing the measure, and by classifying respondents into centrality quintiles for each grade level (from the 20% most popular to the 20% least popular).

Table 1 describes 'popular' teens by regressing centrality on respondent attributes. The results show that popular teens tend to be older, white, from higher income households, well groomed, play sports and they tend to be physically mature and attractive. In contrast, unpopular teens tend to be black or hispanic, overweight, foreign, and new to the school.[7]

## 4.2 Smoking measures

[5]Note that we use the transpose of $A_s$ so that the centrality is proportional to the centrality of the peers who nominate individual $i$, hence, ignoring individuals claiming to be friends with the popular teens.

[6]About 15% of students were not properly matched to the nomination roster and thus their incoming friendship nominations were lost. A similar issue is that friendship nominations are truncated to a maximum of ten nominations. Costenbader and Valente (2003) find that, when working with a subsample of nodes, the eigenvector centrality is the most robust among the eight common centrality measures they consider: across several data sets, eigenvector centralities computed with 50% of the sample have an average correlations higher than 0.95 with eigenvector centralities computed with the full sample.

[7]We find that smoking is not related to popularity; neither is genre.

Table 1: Correlates of popularity –weighted standardized eigenvector centrality.

| | | | | | |
|---|---|---|---|---|---|
| Smoke - Tried 1995 | 0.033 | Black | -0.367*** | Candid (0/1) | 0.003 |
| | (0.028) | | (0.046) | | (0.030) |
| Smoke - Everyday 1995 | -0.022 | Hispanic | -0.123*** | Attractive personality (0/1) | -0.018 |
| | (0.050) | | (0.045) | | (0.038) |
| Cigs. available at home | -0.044 | Asian | 0.014 | Well groomed (0/1) | 0.060* |
| | (0.030) | | (0.068) | | (0.032) |
| Male | -0.002 | Other | -0.13 | Physically attractive (0/1) | 0.182*** |
| | (0.031) | | (0.162) | | (0.037) |
| Age | 0.298* | Weekly earnings | -0.027 | Physically mature (0/1) | 0.118*** |
| | (0.164) | | (0.018) | | (0.032) |
| Age sq. | -0.009* | Overweight (0/1) | -0.151*** | Const. | -2.719* |
| | (0.005) | | (0.028) | | (1.390) |
| HH income | 0.922*** | Sports 1-2 times/week | 0.055 | | |
| | (0.346) | | (0.033) | | |
| Foreign | -0.096** | Sports 3-4 times/week | 0.094** | | |
| | (0.039) | | (0.040) | Degrees of freedom | 148 |
| New student | -0.152*** | Sports 5+ times/week | 0.198*** | $R^2$ | 0.154 |
| | (0.033) | | (0.039) | N | 7169 |

The items on smoking behavior differ slightly between the In-Home and In-School survey instruments. For the In-Home survey, we consider as regular smokers those who smoked *everyday* during the month prior to the interview. We also classify casual smokers as those who *tried* at least one cigarette in the 30 days prior to the interview. More precisely, for each wave we also measure the number of cigarettes consumed per month.[8] In wave 4 (for which the interview took place fourteen years after wave 1) we also code respondents having *ever* smoked everyday for at least one month;[9] the age at which individuals tried their first cigarette; and the age at which individuals started smoking everyday. The top panel of Table 2 summarizes these measures. In 1996, 32% of students tried

Table 2: Summary statistics.

| | Mean | SD | N |
|---|---|---|---|
| ***Smoking*** | | | |
| Tried by 1995 | 0.25 | 0.43 | 7620 |
| Tried 1996 | 0.32 | 0.47 | 7620 |
| Everyday 1996 | 0.11 | 0.31 | 7620 |
| Everyday 2002 | 0.17 | 0.37 | 7620 |
| Everyday 2009 | 0.21 | 0.41 | 6293 |
| Everyday by 2009 | 0.43 | 0.49 | 6290 |
| # cig./month 1996 | 157.01 | 238.05 | 1860 |
| # cig./month 2002 | 302.8 | 260.08 | 1679 |
| # cig./month 2009 | 274.74 | 281.78 | 2179 |
| Age first cigarette | 15.84 | 3.49 | 3955 |
| Age smoked everyday | 17.38 | 3.38 | 2661 |
| ***Peers in 1995*** | | | |
| *Smoking rate among:* | | | |
| 20% most popular | 0.16 | 0.37 | 13394 |
| 80% least popular | 0.16 | 0.37 | 52792 |
| *Mean popularity of:* | | | |
| Smokers | -0.04 | 0.87 | 10657 |
| Non-smokers | 0.01 | 1.02 | 55529 |

Peer smokers are those who smoke at least “once or twice a week”. Popularity is defined by standardized weighted-eigenvector centrality.

[8]Number of cigarettes is computed by multiplying the answers to the questions “During the past 30 days, on how many days did you smoke cigarettes?” and “During the past 30 days, on the days you smoked, how many cigarettes did you smoke each day?”

[9]These measure of smoking should account for individuals who may have smoked only between interviews.

cigarettes and 11% smoked everyday; by 2009, 21% were regular smokers.[10]

For peers in the In-School survey, we define as smokers those who reported smoking "once or twice a week" to "daily". We intentionally use a high threshold of consumption for peers because we suspect that influence arises from regular users who may influence an (initially) soft consumption to the newly initiated individuals, which may eventually result in regular smoking later on.[11] The bottom panel of Table 2 summarizes the smoking patterns of peers classified by their standardized weighted-eigenvector centrality. On average, there is no difference in the smoking propensity of the 20% most popular and the 80% least popular (both have a smoking propensity of 0.16). The standardized popularity of smokers and non-smokers is near zero, indicating that both of these groups have average popularity (see the working paper Robalino (2016) for detailed summary statistics).

## 4.3 Model specification

We test the effects of popularity on peer influence using three related models. With each model, we regress smoking in each of the last three panel waves on peers' smoking in the first panel wave (the 1995 In-School survey) while controlling for school fixed effects.

Equation 1 models the probability of smoking as a probit regression of the smoking behavior $Y^t_{i,s,g}$ of individual $i$ at time $t$ on the mean popularity of smokers, $\bar{P}^{t_0}_{s,g,y=1}$, and the mean popularity of non-smokers, $\bar{P}^{t_0}_{s,g,y=0}$, controlling for the smoking propensity, $\bar{Y}^{t_0}_{s,g}$, at the grade level $g$ in school $s$ during the first wave, $t_0$. Formally we estimate the following equation:

$$Probit(Y^t_{i,s,g} = 1|X_i, s, g) = c + z_s + \beta X_i + \alpha_1 \bar{P}^{t_0}_{s,g,y=0} + \alpha_2 \bar{P}^{t_0}_{s,g,y=1} + \alpha_3 \bar{Y}^{t_0}_{s,g} \quad (1)$$

[10] Our estimation strategy assumes that the correlations between smoking in 1995 and in subsequent years are not so high as to introduce multicollinearity. In Table S1 in the supporting information we present these correlations and the first column shows that there is considerable variation between smoking in 1995 and subsequent years (ranging from a correlation with smoking in 2009 of 0.3, to a correlation with "trying cigarettes in 1996" of 0.53).

[11] All results are similar when we measure peer's smoking as having tried cigarettes; results available upon request.

where $X_i$ is a vector of demographics and household characteristics and $z_s$ is a school fixed effect. Equation 1 formalizes the hypothesized peer effects as $\alpha_2 > 0 > \alpha_1$, that is, the greater the popularity of smokers, the more likely an individual is to smoke in the future; and the greater the popularity of non-smokers, the less likely an individual is to smoke in the future.

Equation 2 models the smoking propensity of the 20% most popular teens, $\bar{Y}^{t_0}_{s,g,Pop}$, and the smoking propensity of the 80% least popular teens, $\bar{Y}^{t_0}_{s,g,noPop}$, in school $s$ and grade $g$:

$$Probit(Y^t_{i,s,g} = 1|X_i, s, g) = c + z_s + \beta X_i + \alpha_1 \bar{Y}^{t_0}_{s,g,Pop} + \alpha_2 \bar{Y}^{t_0}_{s,g,noPop} \qquad (2)$$

Equation 2 formalizes the hypothesized peer effects as $\alpha_1 > \alpha_2$, representing a stronger influence from the popular teens.

Equation 3 models the smoking propensity $\bar{Y}^{t_0}_{s,g,q}$ of the teens in each popularity quintile $q$ in school $s$ and grade $g$:

$$Probit(Y^t_{i,s,g} = 1|X_i, s, g) = c + z_s + X_i\beta + \sum_{q=1}^{q=5} \alpha_q \bar{Y}^{t_0}_{s,g,q} \qquad (3)$$

We hypothesize that $\alpha_5 > \alpha_{q \neq 5}$.

All three models are vulnerable to unobservable confounds, which we address using temporal lags and school fixed effects. Most teens go to their local public school and many parents have little flexibility to switch neighborhoods based only on school choices. Even when they consider the schools in their neighborhoods, they will presumably consider other aspects of the school, including academic performance, class sizes, and facilities, before considering the smoking propensity in the school. We control for new students and for parents who claimed to have chosen their neighborhood in part for the school quality, and we use lagged peers' behavior to model the direction of influence. Most importantly, school fixed effects control for contextual factors such as the local price of cigarettes, the school's rules about smoking, and the local sentiment towards smoking. Thus, our variation is among cohorts within schools, which minimizes the effects of self-selection into schools and contextual confounds.

## 4.4 Instrumental variables

We use instrumental variables to provide additional support for the causal inferences based on temporal lags and school fixed-effects. For model 1 we instrument the popularity of smokers and non-smokers. We need instruments that determine the popularity of peers but do not influence the future smoking choices of others. Based on the correlates of popularity found in Table 1, we instrument peer smokers' mean popularity using the percent of smokers who are new to the school, overweight, physically attractive, physically mature, well groomed, white, and foreign, as well as their mean household income and weekly earnings. We also instrument popularity among non-smokers using these same measures. These measures are only available for the In-Home sample; thus, we use the corresponding averages in the In-Home sample to instrument for the popularity of peers in the In-School sample (see Fig 1). Note that school fixed effects are also included in the first stage regression ensuring that instruments such as percentage of white students and mean household income are uncorrelated with the error term in the second stage regression. The percentage of physically attractive smokers should only affect the smoking choices of others through the resulting prestige and popularity of these smokers.

For model 2 we instrument the smoking propensity among popular and non-popular students. The instruments should affect the smoking choices of peers but not the smoking choices of other individuals. We use the percentage of parents who smoke, percentage of households with smokers, the percentage of parents who are home when school finishes, the percentage of black students, percentage of students with older siblings, the average household income and the average weekly earnings among the 20% most popular and, analogously, among the 80% least popular students. Peers' parents smoking should only affect smoking choices of peers but not of other students.

# 5 Results

Using peers' lagged behavior, school fixed effects, and instrumental variables to control for contextual confounds and school selection, results for model 1 show that the probability that the respondent will take up smoking increases with a lagged measure of the mean popularity of peer smokers as measured by network centrality. Conversely, higher mean popularity of non-smokers decreases this probability. [12] These patterns persist seven and fourteen years after peers' behavior was measured (i.e., in waves 3 and 4 of the data). Similarly, by decomposing the smoking propensity of peers into the propensity of the 20% most popular teens and that of the 80% least popular, results for model 2 show that peer effects are mainly driven by the influence of the 20% most popular teenagers. Furthermore, in the long run we find a negative influence from the smoking propensity of the 80% least popular peers (in waves 3 and 4 of the data). Similar results apply to the number of cigarettes smoked per month as well as to the age of initiation (see Tables A10 and A11 in the supporting information for details). These patterns suggest that social influence on smoking among adolescents is conditioned by the social status of peers: individuals tend to follow the smoking behavior of popular peers and avoid the behavior of unpopular peers.

## 5.1 Popularity of smokers and non-smokers

Table 3 summarizes results for model 1. Consistent with published results from previous studies of peer influence, the third row of the table shows an effect of the aggregate smoking propensity. In 1996, switching from a school grade where no one smokes to one where 25% of students smoke increases the probability of trying cigarettes by 4.8 percentage points and that of smoking every day by 3.98 percentage points, comparable to the effect of having access to cigarettes at home, and similar in magnitude to results reported in previous studies (Fletcher, 2010). Yet,

[12]Patterns are similar when we use alternative popularity measures; see Robalino (2016) for details.

this peer effect vanishes by 2002 and 2009.

Table 3: Probability of smoking and popularity of smokers/non-smokers – probit average marginal effects.

| | Tried 1996 | 1996 | 2002 | 2009 | by 2009 |
|---|---|---|---|---|---|
| Mean popularity of smokers | 0.046*** | 0.009 | 0.027*** | 0.028** | 0.034*** |
| | (0.011) | (0.006) | (0.009) | (0.012) | (0.012) |
| Mean popularity of non-smokers | -0.059*** | -0.014 | -0.034** | -0.031* | -0.062*** |
| | (0.015) | (0.010) | (0.014) | (0.016) | (0.019) |
| % smokers in grade | 0.192* | 0.159** | -0.054 | -0.022 | -0.113 |
| | (0.101) | (0.064) | (0.087) | (0.106) | (0.127) |

Regressions include school fixed effects. Robust standard errors are in parentheses. Peer smokers are those who smoke at least "once or twice a week" in 1995. Peer variables are at the grade level. Includes all covariates from Table S2 in the supporting information.

Our main result is presented in the first two rows of Table 3: the mean popularity of smokers increases the probability of an individual smoking, and the mean popularity of non-smokers decreases this probability. In contrast to the aggregate peer effect, the effect of peers' popularity persists in 2002 and in 2009, suggesting that in the long run the popularity of peer smokers is more important than how many peers smoked. The left panel suggests that an increase of a standard deviation in the mean popularity of smokers results in a increase of 4.6 percentage points in the probability of trying cigarettes in 1996, while the same increase in the mean popularity of non-smokers reduces this probability by 5.9 percentage points. The third panel suggests that an increase of a standard deviation in the popularity of smokers will result in an increase of 2.8 percentage points in the probability of smoking everyday in 2009, fourteen years after having interacted with those peers, and the same variation in the popularity of non-smokers will reduce this probability by 3.1 percentage points. Similarly, the fourth panel suggests that an increase of a standard deviation in the mean popularity of smokers results in an increase of 3.4 percentage points in the probability of ever smoking daily by 2009 and the same increase in the popularity of non-smokers results in a decrease of 6.2 percentage points in the probability of ever smoking.[13]

[13] Note that the probability of smoking *everyday* in 1996 is not significantly affected by the popularity of peers.

**Robustness** One possible explanation for our results is that popular peers may not be more influential due to their social status, but may simply reach more people and that their influence is a simple one, just like that of other peers, but towards more friends. To test this, we include the smoking propensity of all nominated friends for each individual. Table A4 (in the supporting information) summarizes the results. The smoking propensity among direct friends absorbs the effect of the smoking propensity in the grade and it is still significant in 2002 and in 2009 (although there is potential self-selection). In contrast, the effects of the mean popularity of smokers and non-smokers remain essentially unchanged. This is consistent with the hypothesis that popular peers have a stronger influence than non-popular peers.

The persistence of peer effects may be due to addiction, which can be expected to induce serial correlation in smoking measures for latter periods. To control for serial correlation of smoking in the latter waves, we consider separately subsamples of smokers and non-smokers in early waves and predict their probability of starting/quitting smoking in latter waves. The results reported in Table A5 suggest that addiction may eventually override some peer effects in the long term, but not all. The first three columns indicate that individuals who were not regular smokers in 1996, and even those who had never been regular smokers by 2002, continue to be influenced by the popularity of high school smokers/non-smokers to adopt smoking by 2002 or 2009. The last two columns indicate that, among those who smoked in high school (1996), those who were exposed to popular smokers, compared to those who were not exposed, were still more likely to be smoking as late as 2002, but not after. And among those who smoked in 2002, we find no residual effect of the 1995 popularity of smokers by 2009. A reasonable

It seems that influence from peers' popularity only produced casual smoking at that time (i.e. during high school years). Yet, the smoking rate in the grade is significant in this regression. We note that in 1996 the average age in the sample was 16 years old, thus, most smokers could not yet buy cigarettes legally. This may explain why the aggregate smoking propensity was more important for regular smokers in this period as it may have facilitated the supply of cigarettes among peers. If we only consider the subsample of students who had access to cigarettes at home as a proxy of students with weaker supply constraints, the effect from the smoking rate has no statistical significance and we get a significant effect from the popularity of smokers (see Table A3).

conclusion is that by 2009, those who began smoking seven or more years earlier may indeed be hooked and their probability of quitting may be less susceptible to the popularity of smokers and non-smokers in high school.

Table A6 (in the supporting information) reports results of a robustness test using instrumental variables to infer causality. The results confirm those reported in Table 3, although, as expected, we loose some statistical significance. The instruments are highly significant in the first stage regression (J statistics between 13.6 and 22.9 corresponding to p-values of zero for the instruments' joint significance in the first stage regressions); over-identification tests fail to reject the validity of our instruments and Wald test of exogeneity fail to reject the exogeneity of the popularity of smokers and non-smokers' suggesting that a single equation probit model may be more appropriate.[14]

## 5.2 Smoking propensity of popular peers

Table A7 (in the supporting information) reports results estimated from equation 2. The results suggest that peer influence from smokers is driven mainly by the 20% most popular peers.[15] If all of the 20% most popular teens in the grade smoked, the probability of trying cigarettes the following year would increase by 18.5 percentage points (statistically significant at the .01 level), while if all of non-popular teens smoked, resulting in a much larger population of smokers, the increase in this probability would not be statistically different from zero. In fact, in 2002 and 2009, the bottom 80% seem to have a negative influence, that is, the more of these less-popular peers smoke, the less likely an individual is to pick up smoking later in life (although the effect is no longer statistically significant in 2009).

[14]We can reject exogeneity at the 10% level in one regression: having ever smoked regularly by 2009. Results still hold.

[15]We find a similar pattern when considering the 10% most popular and the 90% least popular peers, albeit statistical significance is some what weaker. Results available upon request.

**Robustness** Table A8 (in the supporting information) reports the results for the instrumental regression for Model 2. The results are qualitatively similar, but the peer influence of popular students is not statistically significant. While the model fails the over-identification test only in the 1996 measures, we cannot reject the null hypothesis that the smoking rates of smokers/non-smokers are exogenous, suggesting that a probit model may be more appropriate.

## 5.3 Smoking propensity by popularity quintiles

Table A9 (in the supporting information) reports results for the effects estimated in equation 3. The effects are similar to those in Table A7, although the more granular cut of peer groups makes the pattern less striking. The top popularity quintile is the main driver of influence in every period. We do not find a long-term effect on smoking in 2009, but the probability of having ever smoked by 2009 decreases with the smoking propensity of the second most unpopular quintiles.[16]

## 5.4 Heterogeneity

We tested for possible interaction effects between the influence from smokers' status and individual characteristics (gender, age, age relative to grade mean, physical attractiveness rated by interviewer, a dummy for new students, a dummy for foreigners and a dummy for parents claiming to have chosen the neighborhood partly for schools' quality), as well as individual's own network variables (the popularity of the respondent; a dummy signaling whether the respondent's best friend also nominated him/her as a best friend; and the overall percentage of reciprocal ties from the individual's nominations), and school social network characteristics (network density of the school). No statistically significant interaction effects were observed,

[16]This may be because the lowest popularity quintile may have been disconnected from the social scene and, thus, forgotten; while the second lowest popularity quintile may have been more connected to the social scene while retaining the lowest social status among the connected students.

suggesting that all students are vulnerable to the influence of popular smokers.[17]

# 6 Conclusion

There is a general consensus in the literature that adolescents are susceptible to peer influences, to engage in risky behaviors, including smoking. Previous studies have identified the desire for social status as an important motivation. The clear implication is that influence on smoking behavior from popular peers is stronger than influence from unpopular peers. We test this hypothesis using four waves of panel data from AddHealth. These data are unusual in providing a nearly complete social network of the school population, with which to measure popularity as the respondent's eigenvector centrality in the network. An important limitation of these data is "the reflection problem." We address challenges to causal inference posed by selection and contextual effects using school fixed effects, lagged measures of smoking adoption, and instrumental variables.

The results show that the probability of smoking increases with the lagged popularity of smokers, while the popularity of non-smokers has the opposite effect. These effects persist up to fourteen years after exposure to peers' behavior. Drilling down, we find that most of the aggregate peer effects come from the smoking propensity of the 20% most popular peers. Importantly, we also find evidence of negative social influence, in which respondents avoid the behavior of unpopular teens. The smoking propensity of the 80% least popular peers has a negative influence on individuals' smoking in the long run (seven and fourteen years later).

These findings underscore the importance of the popularity of peers in predicting their influence. Not only is it important to know the smoking propensity in schools but also where the smokers are located in the social hierarchy. For example, our results suggest that two schools

[17] For continuous variables, we regressed separately continuous interactions as well as a dummy signaling values above and below the median. See Robalino (2016) for details. All results available upon request.

where 20% of teens smoke may experience very different trajectories in the incidence of smoking: If those 20% who smoke are the most popular teens, the probability of smoking for students the next year can be expected to increase by 18 percentage points, while if they are among the least popular, the probability would not increase by a statistically significant amount.[18]

These results point to several possible extensions. Although status has been shown to condition peer effects on the diffusion of a number of social contagions, more research is needed to know if popularity conditions peer influence on a range of other risky behaviors, including binge drinking, drug abuse, and stigmatized sexual behaviors. Equally important, we need to know if these status effects extend as well to positive adolescent behaviors such as academic achievement and sports.

[18] Collecting the entire social network of schools to compute popularity with a network centrality may be costly; yet, Banerjee et al. (2014) find that asking a subsample of individuals who are the influential people can lead to accurate identification of central people.

## References

Alexander, C., M. Piazza, D. Mekos, and T. Valente (2001). Peers, Schools, and Adolescent Cigarette Smoking. *Journal of Adolescent Health 29*(1), 22–30.

Aral, S. and D. Walker (2012, July). Identifying Influential and Susceptible Members of Social Networks. *Science 337*(6092), 337–341.

Banerjee, A. V., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2014, June). Gossip: Identifying Central Individuals in a Social Network. *arXiv.org*.

Campbell, R., F. Starkey, J. Holliday, S. Audrey, M. Bloor, N. Parry-Langdon, R. Hughes, and L. Moore (2008, May). An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial. *The Lancet 371*(9624), 1595–1602.

Christakis, N. A. and J. H. Fowler (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine 357*(4), 370–379.

Christakis, N. A. and J. H. Fowler (2008). The Collective Dynamics of Smoking in a Large Social Network. *New England Journal of Medicine 358*(21), 2249–2258.

Clark, A. E. and Y. Lohéac (2007, July). "It Wasn't Me, It Was Them!" Social Influence in Risky Behavior by Adolescents. *Journal of Health Economics 26*(4), 763–784.

Cohen-Cole, E. and J. M. Fletcher (2008, December). Detecting Implausible Social Network Effects in Acne, Height, and Headaches: Longitudinal Analysis. *BMJ 337*(dec04 2), a2533–a2533.

Costenbader, E. and T. W. Valente (2003). The Stability of Centrality Measures When Networks Are Sampled. *Social Networks 25*(4), 283–307.

Ennett, S. T., K. E. Bauman, A. Hussong, R. Faris, V. A. Foshee, L. Cai, and R. H. DuRant (2006). The Peer Context of Adolescent Substance Use: Findings from Social Network Analysis. *Journal of Research on Adolescence 16*(2), 159–186.

Everett, S. A., C. W. Warren, D. Sharp, L. Kann, C. G. Husten, and L. S. Crossett (1999, November). Initiation of Cigarette Smoking and Subsequent Smoking Behavior among U.S. High School Students. *Preventive Medicine 29*(5), 327–333.

Fletcher, J. M. (2010, April). Social Interactions and Smoking: Evidence Using Multiple Student Cohorts, Instrumental Variables, and School Fixed Effects. *Health Economics 19*(4), 466–484.

Gardner, M. and L. Steinberg (2005, July). Peer Influence on Risk Taking, Risk Preference, and Risky Decision Making in Adolescence and Adulthood: an Experimental Study. *Developmental Psychology 41*(4), 625–635.

Gaviria, A. and S. Raphael (2001). School-Based Peer Effects and Juvenile Behavior. *Review of Economics and Statistics 83*(2), 257–268.

Gladwell, M. (2006). *The Tipping Point: How Little Things Can Make a Big Difference*. Little, Brown.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton, NJ : Princeton University Press.

Krauth, B. V. (2006, July). Simulation-Based Estimation of Peer Effects. *Journal of Econometrics 133*(1), 243–271.

Lando, H. A., D. T. Thai, D. M. Murray, L. A. Robinson, R. W. Jeffery, N. E. Sherwood, and D. J. Hennrikus (1999, December). Age of Initiation, Smoking Patterns, and Risk in a Population of Working Adults. *Preventive Medicine 29*(6 Pt 1), 590–598.

Lundborg, P. (2006, March). Having the Wrong Friends? Peer Effects in Adolescent Substance Use. *Journal of Health Economics 25*(2), 214–233.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies 60*(3), 531.

Mercken, L., T. A. B. Snijders, C. Steglich, E. Vartiainen, and H. de Vries (2010, January). Dynamics of Adolescent Friendship Networks and Smoking Behavior. *Social Networks 32*(1), 72–81.

Michell, L. and A. Amos (1997). Girls, Pecking Order and Smoking. *Social Science & Medicine 44*(12), 1861–1869.

Nakajima, R. (2007). Measuring Peer Effects on Youth Smoking Behaviour. *Review of Economic Studies 74*(3), 897–935.

Norton, E. C., R. C. Lindrooth, and S. T. Ennett (1998). Controlling for the Endogeneity of Peer Substance Use on Adolescent Alcohol and Tobacco Use. *Health Economics 7*(5), 439–453.

Oldmeadow, J. A., M. J. Platow, M. Foddy, and D. Anderson (2003). Self-categorization, status, and social influence. *Social Psychology Quarterly*, 138–152.

Powell, L. M., J. A. Tauras, and H. Ross (2005, September). The Importance of Peer Effects, Cigarette Prices and Tobacco Control Policies for Youth Smoking Behavior. *Journal of Health Economics 24*(5), 950–968.

Robalino, J. D. (2016). Smoking peer effects among adolescents: Are popular teens more influential? *IZA DP* (9714).

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

Steglich, C. and T. Snijders (2010). Dynamic Networks and Behavior: Separating Selection from Influence. *Sociological Methodology 40*(1), 329–393.

Valente, T. W., J. B. Unger, and C. A. Johnson (2005, October). Do Popular Students Smoke? The Association between Popularity and Smoking among Middle School Students. *Journal of Adolescent Health 37*(4), 323–329.

# Appendix

Table A1: Correlation across measures of smoking.

| | Tried by 1995 | Tried 1996 | 1996 | 2002 | 2009 |
|---|---|---|---|---|---|
| Tried by 1995 | 1 | | | | |
| Tried 1996 | 0.535 | 1 | | | |
| 1996 | 0.492 | 0.523 | 1 | | |
| 2002 | 0.349 | 0.391 | 0.427 | 1 | |
| 2009 | 0.305 | 0.338 | 0.367 | 0.572 | 1 |
| by 2009 | 0.43 | 0.488 | 0.399 | 0.56 | 0.59 |

Table A2: Probability of smoking – probit average marginal effects.

| | Tried 1996 | 1996 | 2002 | 2009 | by 2009 |
|---|---|---|---|---|---|
| Mean popularity of smokers | 0.046*** | 0.009 | 0.027*** | 0.028** | 0.034*** |
| | (0.011) | (0.006) | (0.009) | (0.012) | (0.012) |
| Mean popularity of non-smokers | -0.059*** | -0.014 | -0.034** | -0.031* | -0.062*** |
| | (0.015) | (0.010) | (0.014) | (0.016) | (0.019) |
| % smokers in grade | 0.192* | 0.159** | -0.054 | -0.022 | -0.113 |
| | (0.101) | (0.064) | (0.087) | (0.106) | (0.127) |
| Had tried smoking in 1995 | 0.383*** | 0.184*** | 0.152*** | 0.207*** | 0.388*** |
| | (0.007) | (0.005) | (0.008) | (0.010) | (0.011) |
| Male | -0.003 | 0.001 | 0.034*** | 0.050*** | 0.072*** |
| | (0.009) | (0.006) | (0.008) | (0.010) | (0.011) |
| Age (+ Age squ.) | 0.001 | 0.002 | -0.013*** | -0.008* | -0.015** |
| | (0.005) | (0.003) | (0.004) | (0.005) | (0.006) |
| White | *Ommitted* | | | | |
| Black | -0.108*** | -0.081*** | -0.099*** | -0.034** | -0.066*** |
| | *(0.015)* | (0.009) | (0.013) | (0.017) | (0.019) |
| Hispanic | 0.001 | -0.018 | -0.065*** | -0.068*** | -0.02 |
| | (0.019) | (0.014) | (0.016) | (0.018) | (0.022) |
| Asian | -0.024 | -0.03 | -0.025 | -0.034 | 0.021 |
| | (0.026) | (0.019) | (0.026) | (0.030) | (0.033) |
| Other | 0.022 | 0.054 | -0.080** | 0.01 | 0.014 |
| | (0.044) | (0.038) | (0.032) | (0.049) | (0.052) |
| Foreign | -0.02 | -0.02 | -0.006 | -0.028 | -0.041 |
| | (0.021) | (0.019) | (0.022) | (0.027) | (0.028) |
| Out of school in 1996 | 0.061*** | 0.063*** | 0.070*** | 0.060*** | 0.081*** |
| | (0.020) | (0.011) | (0.018) | (0.022) | (0.028) |
| New student in 1995 | 0.005 | 0.003 | -0.01 | 0.019 | 0.024* |
| | (0.011) | (0.008) | (0.010) | (0.012) | (0.014) |
| Weekly earnings ($100) | 0.009* | 0.012*** | 0.001 | 0.005 | 0.015** |
| | (0.006) | (0.004) | (0.005) | (0.006) | (0.007) |
| Household Income ($1000 000) | -0.052 | -0.226* | -0.278** | -0.223 | -0.307** |
| | (0.096) | (0.129) | (0.133) | (0.152) | (0.133) |
| Moved partly for school quality | -0.011 | -0.014** | -0.014* | -0.018* | -0.012 |
| | (0.009) | (0.006) | (0.008) | (0.010) | (0.012) |
| Mother smokes | 0.009 | 0.027*** | 0.015 | 0.031*** | 0.034** |
| | (0.011) | (0.007) | (0.010) | (0.012) | (0.014) |
| Father smokes | 0.023** | 0.016** | 0.035*** | 0.015 | 0.009 |
| | (0.011) | (0.007) | (0.009) | (0.011) | (0.014) |
| Cigarettes at home | 0.045*** | 0.043*** | 0.032*** | 0.054*** | 0.069*** |
| | (0.012) | (0.007) | (0.010) | (0.012) | (0.014) |
| Chi sq. | 2123.537 | 1313.609 | 932.641 | 887.763 | 1386.09 |
| Degrees of freedom | 138 | 123 | 138 | 136 | 140 |
| Pseudo R sq. | 0.257 | 0.361 | 0.151 | 0.153 | 0.191 |
| N | 7630 | 7169 | 7599 | 6256 | 6305 |

Regressions include school fixed effects. Robust standard errors are in parenthesis. Peer smokers are those who smoke at least “once or twice a week” in 1995.

Table A3: Probability of smoking every day in 1996 for students with and without access to cigarettes at home – probit average marginal effects.

| | Access to cigarettes at home | |
|---|---|---|
| | Yes | No |
| Mean popularity smokers | 0.038** | 0.002 |
| | (0.016) | (0.007) |
| Mean popularity non-smokers | -0.021 | -0.023* |
| | (0.024) | (0.012) |
| % smokers | 0.117 | 0.196*** |
| | (0.141) | (0.070) |
| N | 2101 | 4633 |

Regressions include school fixed effects. Robust standard errors are in parenthesis. Peer smokers are those who smoke at least "once or twice a week" in 1995. Peer variables are at the grade level. Includes all covariates from Table S2.

Table A4: Probability of smoking controlling for smoking of direct friends – probit average marginal effects.

| | Tried 1996 | 1996 | 2002 | 2009 | by 2009 |
|---|---|---|---|---|---|
| Mean popularity smokers | 0.052*** | 0.007 | 0.029*** | 0.025** | 0.038*** |
| | (0.012) | (0.006) | (0.010) | (0.013) | (0.013) |
| Mean popularity non-smokers | -0.065*** | -0.014 | -0.033** | -0.029* | -0.056*** |
| | (0.016) | (0.011) | (0.015) | (0.017) | (0.020) |
| % Smokers | 0.117 | 0.081 | -0.157* | -0.106 | -0.237* |
| | (0.104) | (0.066) | (0.092) | (0.112) | (0.132) |
| % Direct friends smoking | 0.119*** | 0.098*** | 0.113*** | 0.130*** | 0.176*** |
| | (0.021) | (0.011) | (0.018) | (0.021) | (0.027) |

Regressions include school fixed effects. Robust standard errors are in parenthesis. Peer smokers are those who smoke at least "once or twice a week" in 1995. Peer variables are at the grade level. Includes all covariates from Table S2.

Table A5: Subsamples of smokers/non-smokers to test for serial correlation in smoking measures – probit average marginal effects.

| Subsample – regular smoker in 1996: | No | | | Yes | |
|---|---|---|---|---|---|
| | 2002 | 2009 | by 2009 | 2002 | 2009 |
| Mean pop. Smokers | 0.015* | 0.02 | 0.031** | 0.352*** | 0.073 |
| | (0.008) | (0.013) | (0.013) | (0.099) | (0.114) |
| Mean pop. Non-Smokers | -0.014 | -0.030* | -0.059*** | -0.322*** | -0.001 |
| | (0.013) | (0.016) | (0.021) | (0.085) | (0.092) |
| % Smokers | -0.093 | -0.087 | -0.218 | -0.166 | -0.043 |
| | (0.091) | (0.115) | (0.142) | (0.357) | (0.381) |
| N | 6714 | 5593 | 5638 | 765 | 603 |

| Subsample – regular smoker *by* 2002: | No | | Yes |
|---|---|---|---|
| | 2009 | by 2009 | 2009 |
| Mean pop. Smokers | 0.017 | 0.001 | 0.035 |
| | (0.016) | (0.012) | (0.027) |
| Mean pop. Non-Smokers | -0.044*** | -0.052*** | -0.019 |
| | (0.015) | (0.019) | (0.038) |
| % Smokers | -0.228** | -0.339** | 0.212 |
| | (0.108) | (0.142) | (0.230) |
| N | 2551 | 3216 | 2042 |

Regressions include school fixed effects. Robust standard errors are in parenthesis. Peer smokers are those who smoke at least “once or twice a week” in 1995. Peer variables are at the grade level. Includes all covariates from Table S2.

Table A6: Instrumental variable probit regression model 1: popularity of smokers and non-smokers.

| | Tried 1996 | 1996 | 2002 | 2009 | by 2009 |
|---|---|---|---|---|---|
| *Probit Coef.* | | | | | |
| Mean popularity smokers | 0.132 | 0.307 | 0.579*** | 0.12 | 0.343* |
| | (0.198) | (0.298) | (0.224) | (0.203) | (0.202) |
| Mean popularity non-Smokers | -0.577** | -0.343 | -0.365 | -0.322 | -0.661** |
| | (0.243) | (0.310) | (0.260) | (0.287) | (0.260) |
| *Marginal effect* | | | | | |
| Mean popularity smokers | 0.034 | 0.04 | 0.124 | 0.029 | 0.105 |
| Mean popularity non-smokers | -0.15 | -0.044 | -0.078 | -0.079 | -0.203 |
| p-value Wald exogeneity test | 0.268 | 0.297 | 0.113 | 0.699 | 0.078 |
| J statistic | 13.63 | 22.899 | 18.496 | 14.646 | 19.267 |
| p-value J statistic (over-identification) | 0.626 | 0.116 | 0.296 | 0.551 | 0.255 |
| F statistic (Mean pop. Smokers) | 40.39 | 40.39 | 40.639 | 34.628 | 34.181 |
| p-value F statistic (Mean pop. Smokers) | 0 | 0 | 0 | 0 | 0 |
| F statistic (Mean pop. Non-Smokers) | 30.587 | 30.587 | 30.471 | 23.186 | 23.48 |
| p-value F statistic (Mean pop. Non-Smokers) | 0 | 0 | 0 | 0 | 0 |
| N | 7089 | 6688 | 7012 | 5807 | 5841 |

Instrumental variables: percentage of white, foreigners, new students, overweight students, physically attractive, physically mature, well groomed, mean household income, and weekly earnings. IVs are computed both among smokers and among non-smokers. These variables are only available in the InHome survey, hence, we compute them from the InHome sample to instrument the mean popularity in the InSchool sample (see Fig 1). Regressions include school fixed effects. Robust standard errors are in parenthesis. Peer smokers are those who smoke at least “once or twice a week” in 1995. Peer variables are at the grade level. Includes all covariates from Table S2.

Table A7: Probability of smoking and smoking rates among popular and non-popular students – probit average marginal effects.

| | Tried 1996 | 1996 | 2002 | 2009 | by 2009 |
|---|---|---|---|---|---|
| *Smoking propensity:* | | | | | |
| 20% most popular | 0.185*** | 0.060* | 0.174*** | 0.031 | 0.157** |
| | (0.052) | (0.031) | (0.045) | (0.056) | (0.065) |
| 80% least popular | 0.033 | 0.100* | -0.224*** | -0.051 | -0.327*** |
| | (0.092) | (0.058) | (0.080) | (0.097) | (0.116) |

Regressions include school fixed effects. Robust standard errors are in parenthesis. Peer smokers are those who smoke at least "once or twice a week" in 1995. Peer variables are at the grade level. Includes all covariates from Table S2.

Table A8: Instrumental variable probit regression model 2: smoking rate among popular and non-popular.

| | Tried 1996 | 1996 | 2002 | 2009 | by 2009 |
|---|---|---|---|---|---|
| *Probit Coef.* | | | | | |
| smoking rate amon 20% most pop. | 0.908 | 0.482 | 0.638 | 1.247 | 0.753 |
| | (0.832) | (1.090) | (0.908) | (0.879) | (0.799) |
| smoking rate amon 80% lest pop. | -0.299 | 0.086 | -0.804 | 0.118 | 0.163 |
| | (0.901) | (1.137) | (0.948) | (1.010) | (0.920) |
| *Marginal effect* | | | | | |
| smoking rate amon 20% most pop. | 0.236 | 0.06 | 0.135 | 0.303 | 0.235 |
| smoking rate amon 80% lest pop. | -0.078 | 0.011 | -0.171 | 0.029 | 0.051 |
| p-value Wald exogeneity test | 0.875 | 0.728 | 0.946 | 0.231 | 0.227 |
| J statistic | 26.182 | 24.051 | 15.824 | 11.891 | 17.411 |
| p-value J statistic (over-identification) | 0.025 | 0.045 | 0.324 | 0.615 | 0.235 |
| F statistic (sm. rate 20% most pop.) | 35.375 | 35.375 | 35.126 | 32.778 | 32.775 |
| p-value F statistic (sm. rate 20% most pop.) | 0 | 0 | 0 | 0 | 0 |
| F statistic (sm. rate 80% least pop.) | 105.947 | 105.947 | 105.114 | 89.985 | 89.948 |
| p-value F statistic (sm. rate 80% least pop.) | 0 | 0 | 0 | 0 | 0 |
| N | 7789 | 7252 | 7756 | 6400 | 6447 |

Instrumental variables: % black students, % of students with older siblings, percentage of parents who smoke, % of households with smokers, % of parents home after school finishes, average household income and average weekly earnings. IVs are computed both among the 20% most popular and the 80% least popular students. These variables are only available in the InHome survey, hence, we compute them from the InHome sample to instrument the mean popularity in the InSchool sample (see Fig 1). Regressions include school fixed effects. Robust standard errors are in parenthesis. Peer smokers are those who smoke at least "once or twice a week" in 1995. Peer variables are at the grade level. Includes all covariates from Table S2.

Table A9: Probability of smoking and smoking propensity by popularity quintiles of peers – probit average marginal effects.

| | Tried 1996 | 1996 | 2002 | 2009 | by 2009 |
|---|---|---|---|---|---|
| *Smoking propensity:* | | | | | |
| Top pop. quintile | 0.218*** | 0.039* | 0.181*** | 0.029 | 0.185** |
| | (0.070) | (0.022) | (0.045) | (0.060) | (0.083) |
| 2nd pop. quintile | 0.091 | 0.025 | -0.087* | -0.024 | -0.089 |
| | (0.081) | (0.026) | (0.051) | (0.065) | (0.096) |
| 3rd pop. quintile | 0.094 | 0.042 | -0.027 | 0.108 | 0.09 |
| | (0.081) | (0.027) | (0.054) | (0.070) | (0.097) |
| 4th pop. quintile | -0.12 | -0.004 | -0.03 | -0.071 | -0.310*** |
| | (0.075) | (0.025) | (0.049) | (0.063) | (0.089) |
| 5th pop. quintile | -0.025 | 0.004 | -0.074 | -0.037 | -0.069 |
| | (0.071) | (0.023) | (0.048) | (0.060) | (0.085) |

Regressions include school fixed effects. Robust standard errors are in parenthesis. Peer smokers are those who smoke at least “once or twice a week” in 1995. Peer variables are at the grade level. Includes all covariates from Table S2.

## 6.1 Number of cigarettes per month and age of initiation

In Table A10 we regress the number of cigarettes consumed per month, as well as the age at which individuals tried their first cigarette and the age at which they smoked everyday for the first time. To account for non-smokers, we estimate a Tobit censured equation analogous to equation (1). We find that the same patterns reported above regarding peers' popularity apply to the number of cigarettes and the age of initiation. Conditional on smoking, a standard deviation increase in the mean popularity of smokers increases consumption by 7.3 cigarettes per month in 1996, 12.5 cigarettes per month in 2002 and 8.7 per month in 2009; and the same variation in the popularity of non-smokers decreases consumption by 9.2 cigarettes per month in 1996, 18.6 cigarettes per month in 2002 and 20.4 in 2009.

Table A10: Number of cigarettes smoked per month and age of initiation – Tobit average marginal effects.

| | Number of cigarettes/ month | | | Age | |
|---|---|---|---|---|---|
| | 1996 | 2002 | 2009 | 1st cig. | smoked everyday |
| Mean popularity smokers | 7.302*** | 12.490*** | 8.739** | -0.824*** | -0.663*** |
| | (2.295) | (4.678) | (4.339) | (0.243) | (0.254) |
| Mean popularity non-smokers | -9.167** | -18.586** | -20.447*** | 1.044** | 1.155*** |
| | (3.779) | (7.432) | (7.453) | (0.426) | (0.372) |
| % smokers | 84.879*** | -8.14 | -62.182 | 1.521 | 2.634 |
| | (29.591) | (46.456) | (48.517) | (2.625) | (2.282) |
| N | 7605 | 6119 | 6266 | 6284 | 6301 |

Regressions include school fixed effects. Robust standard errors are in parentheses. Peer smokers are those who smoke at least "once or twice a week" in 1995. Peer variables are at the grade level. Includes all covariates from Table S2 in the supporting information.

Regarding the age of initiation, conditional on ever trying cigarettes, we find that a standard deviation above the mean popularity of smokers advances the age at which individuals try their first cigarette by 10 months; and the same variation in the popularity of non-smokers delays the age at which individuals try their first cigarette by over a year. Similarly, conditional on ever smoking everyday, we find that a standard deviation above the mean popularity of smokers advances the age at which individuals started smoking everyday by almost 8 months; and the same variation in the popularity of non-smokers delays the age at which individuals started smoking everyday by over a year.

Table A11: Number of cigarettes smoked per month and age of initiation – model 2 Tobit average marginal effects.

| | Number of cigarettes/ month | | | Age | |
|---|---|---|---|---|---|
| | 1996 | 2002 | 2009 | 1st cig. | smoked everyday |
| *Smoking propensity:* | | | | | |
| 20% most popular | 27.888** | 45.655** | 1.119 | -2.970** | -2.567** |
| | (11.720) | (23.232) | (24.535) | (1.437) | (1.199) |
| 80% least popular | 56.649** | -51.03 | -63.971 | 3.717 | 5.887*** |
| | (26.144) | (42.433) | (42.551) | (2.461) | (2.075) |

Regressions include school fixed effects. Robust standard errors are in parenthesis. Peer smokers are those who smoke at least "once or twice a week" in 1995. Peer variables are at the grade level. Includes all covariates from Table S2.

Table A11 reports estimates for Tobit regressions analogous to model 2. If all of the 20% most popular students smoked, cigarettes consumption would increase by 27.9 cigarettes in 1996, and 45.6 in 2002 (but no statistical significance for 2009); the same variation advances the age of the first cigarette by almost three years and the age when an individual first started smoking everyday by 2.6 years; if 25% of the 80% least popular teens smoked (an equivalent 20% of the total students), cigarettes consumption would increase by 14 cigarettes in 1996 (no significance in 2002 or 2009) and delays the age when first started smoking everyday by almost 1.5 years.[19]

These patterns are of particular importance in light of the existing evidence that smoking at a younger age has a strong impact in the number of cigarettes smoked and the probability of quitting smoking later in life (Everett et al., 1999; Lando et al., 1999). Popularity of peer smokers during the teenage years seems to be an important driver of these findings.

[19] We find similar patterns using peers' smoking propensity by popularity quintiles; results available upon request.

# High School GPA and Popularity: Does the Popularity of High-Achievers Improves Individuals' performance?

Juan David Robalino[†*]
[†]Department of Economics, Cornell University
[†]IZA – Institute for the Study of Labor
[*]E-mail: jr872@cornell.edu

**Abstract**

We analyze peer effects in education by considering the 'popularity' of good and bad students. The analysis is based on a four wave panel survey representative of American high school students. We measure peers' popularity by their eigenvector centrality in high school social networks. We rely on variations across adjacent cohorts within schools by using school fixed effects, lagged peers' behavior, and instrumental variables to control for homophily and contextual confounds. We find that the popularity of peers with good grades in mathematics increases considerably individuals' performance the following year, while the popularity of bad students has the opposite effect. The positive effect is somewhat stronger for males, while the negative effect is much stronger for females. Results suggest that peer effects in education may not only work through peers' performance but may also be mediated by the social status of high and low achievers.

"It's cool to be smart." – President Barack Obama, Apr. 29, 2009.

# 1 Introduction

Education is one of the main factors influencing economic opportunities. Yet, an important share of our investments in education take place during adolescence, a tumultuous life period when individuals want to fit in and are particularly vulnerable to peer pressure (Brown 2004). Since the seminal work by James Coleman (1961), it has been acknowledged that adolescents face trade-offs between peers' approval and academic goals. Negative stigmas towards high-achievers, such as labelings of "geeks" and other forms of bulling can discourage academic performance. For example, the influential work by Fordham and Ogbu (1986) developed the "acting white" hypothesis, suggesting that black students may be discouraged to excel academically in order to avoid judgement from their peers. Subsequently, Tyson and Darity Jr (2005) showed that high-achieving students are typically stigmatize as "nerds" regardless of race.

In this paper we explore how the popularity (i.e., social network centrality measures) of high-achieving peers in high school may mediate individuals' academic performance. The hypothesis is that popular individuals are likely to implicitly set social norms in schools, thus their academic performance can influence academic goals of other students. Similarly, if high-achieving students are popular, students may be spared the potential stigma associated with academic excellence.

The study is based on data from the National Longitudinal Survey of Adolescent Health (AddHealth), a four-wave panel (1995, 1996, 2002 and 2009) representative of American High Schools. The survey contains exhaustive data on social networks that we use to measure the popularity of students through network centrality measures. To overcome self-selection and contextual confounds, we exploit variations in peers' popularity and GPA across adjacent cohorts within schools.

We find that individuals' GPA in mathematics increases considerably with the popularity of peers who have a GPA above average, and decreases by a similar magnitude with the popularity of peers with GPA bellow average. These results are robust across model specifications including when using instrumental variables. The popularity and academic performance of peers also seems to influence the probability of attending and completing college.

The reminder of this paper is organized as follows: Section 2 discusses the existing work on peer effects in education. Section 3 discusses the data and our identification strategy. Section 4 presents the results, and Section 5 concludes.

# 2 Background

The seminal work by Manski (1993) uncovered many difficulties in the estimation of peer effects, which he referred to as *the reflection problem*. One important issue is that a correlation between the behavior of the group and an individual's behavior can be attributed to three possible mechanisms:

- *endogenous effects* where an individual's choice is influenced by the choices of the group;
- *exogenous (contextual) effects* where individuals in a given group may behave similarly because the whole group has experienced an (unobserved) exogenous shock;
- *correlated effects* where individuals in a group behave similarly because they have similar unobservable characteristics and self-select into the group.

Peer effects are an endogenous mechanism. However, in observational data it is often difficult to disentangle endogenous effects from contextual and correlated effects.

There is a broad literature on peer effects in education using two main identification strategies. Sacerdote (2001) set the gold standard looking at peer effects among college students. His approach consists of using exogenous group formation from random allocation to college dorms in order to control for peers' self-selection and homophily, and peers' lagged GPA (from high-school) to control for contextual confounds. Several other studies have also found peer effects following this strategy (e.g., Stinebrickner and Stinebrickner, 2006; Zimmerman, 2003). In particular, Carrell and Fullerton (2009) found very strong peer effects using random allocation to "squadrons" in the U.S. Air Force Academy –a particularly tight peer group.

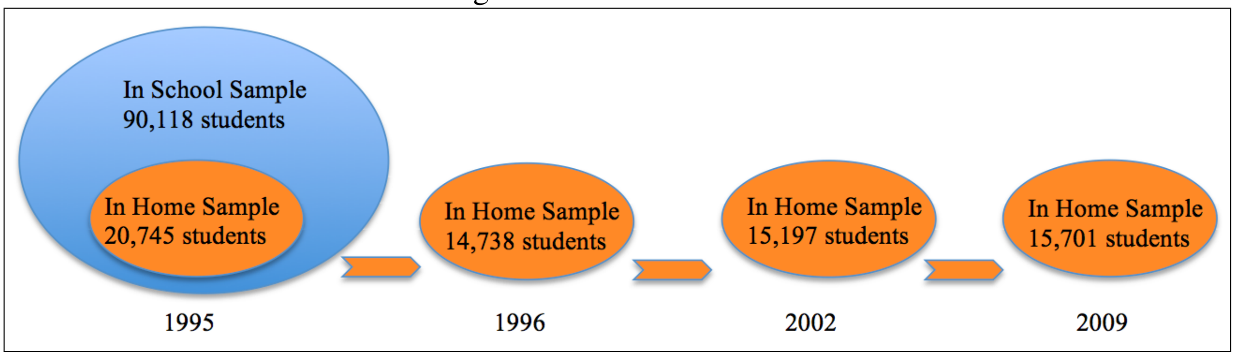
Yet, exogenous peer group formations are not common in secondary education. Following Hoxby (2000), the second main identification strategy has been to use idiosyncratic variations in the composition of peer groups across adjacent cohorts within schools by using school fixed effects. This strategy relies on the assumption that parents may sort their children based on school characteristics, but not based on cohort characteristics. Hoxby (2000) used gender variations across cohorts as an unexpected shock to peers' performance –since females tend to be better students in average– and she found peer effects in reading scores and particularly large effects in mathematics. Several other papers followed this approach such as Hanushek et al. (2003) who used cohort variations in peers' past performance, McEwan (2003) and Bifulco et al. (2011) who used the education of peers' mothers, and Lavy et al. (2011) who used the percentage of peers who repeated years. All these papers found significant peer effects. We will follow a similar strategy as described in detail in the next section.

A particularly relevant paper for the current study is that of Bursztyn and Jensen (2015). They ran a field experiment where they offered complementary SAT prep classes to high school students in both honours and non-honours classes, either under a private condition or a public condition where peers would know who signed-in. They found that the public condition decreased students uptake when exposed to their non-honours classes peers, but the public condition increased uptake when exposed to their honours class peers. Furthermore, for students who claimed that it is important to be popular, the sign-up gap when exposed to honours vs.

non-honours was even wider, and there was no significant effect for those who claimed it is not important to be popular. Their results suggest that students may adapt their behavior to comply with the implicit social norm of peers (e.g., an academically oriented norm in honours classes).

# 3 Data and Identification Strategy

Figure 1: Data structure.




## 3.1 Data

The present study is based on data from the National Longitudinal Survey of Adolescent Health (AddHealth), a four-wave panel (1995, 1996, 2002 and 2009) administered to a representative sample of American High Schools. This data set contains students' GPA by subjects as well as exhaustive data on social networks that we use to measure the popularity of students. The multi-wave longitudinal data also includes identifiers for grade, year, school, and detailed measures of household and parents' characteristics. The In-School survey conducted in 1995 covered 90,118 adolescents in 144 schools. A representative sub-sample of 20,745 students also completed the 1995 In-Home wave 1 survey which included more detailed information such as family income and parents' education. Finally, about 15,000 students were surveyed again in 1996 (In-Home survey wave 2), as well as 2002 (In-Home wave 3) and 2009 (In-Home wave 4; see Fig 1). After excluding individuals with missing information and schools with insufficient social network data, we have a sample of about 60,000 peers and a core sample of about 7,000 individuals.

## 3.2 Popularity Measure

To identify the popularity of peers we use the In-School friendship nomination. Each participant was asked to identify their five closest male friends and their five closest female friends in order of importance (i.e., best friend, second best friend, etc). This allows us to compute network centrality measures for individuals within each school.

In particular, we consider the *eigenvector centrality*:[1] This centrality measure is of particular interest as it captures the status of an individual. The intuition is that the eigenvector centrality of individual $i$ is proportional to the centrality of peers who nominated $i$ as a friend, hence being nominated by central people makes you more central. In terms of adolescents, we define individual $i$ as popular when the other popular teens consider him/her a friend. Formally, the centrality $v_i$ of individual $i$ in school $s$ is:

$$v_i = \frac{1}{\lambda} \sum_j v_j a_{ji}$$

where $a_{ij} = 1/n_{ij}$ if individual $i$ nominated individual $j$ as his/her $n^{th}$ best friend and $a_{ij} = 0$ otherwise. Equivalently, using vector notation, we recover the eigenvector expression $\lambda v = A_s' v$, where $A_s$ is the so called adjacency matrix of the social network in school $s$ and $\lambda$ is the largest eigenvalue –in order to ensure that all the elements of $v$ are positive as guaranteed by the Perron–Frobenius theorem.[2] This weighted-eigenvector centrality (i.e., weighted by the inverse of the friendship nomination order) is our preferred popularity measure as it also takes account of the difference between being the best friend of many people and being the fifth best

[1] We also consider in-degree and Katz–Bonacich centralities and we obtain similar results (available upon request); see Jackson (2008) Chapter 2 for a discussion about network centrality measures.

[2] Note that we use the transpose of $A_s$ so that the centrality is proportional to the centrality of the peers who nominate individual $i$, hence, ignoring individuals claiming to be friends with the popular teens.

friend of many people.[3] We use a standardized version of our popularity measure for ease of interpretation.

Table 1: Popularity – weighted standardized eigenvector centrality – and its correlates

| | | | | | |
|---|---|---|---|---|---|
| Math GPA | 0.024 | New student | -0.133*** | Candid (0/1) | -0.007 |
| | (0.019) | | (0.049) | | (0.039) |
| History GPA | 0.019 | Black | -0.364*** | Attractive personality (0/1) | -0.043 |
| | (0.021) | | (0.060) | | (0.046) |
| English GPA | 0.052** | Hispanic | -0.084 | Well groomed (0/1) | 0.049 |
| | (0.022) | | (0.058) | | (0.041) |
| Science GPA | 0.005 | Asian | -0.109 | Physically attractive (0/1) | 0.231*** |
| | (0.020) | | (0.085) | | (0.048) |
| Male | 0.027 | Other | -0.114 | Physically mature (0/1) | 0.110*** |
| | (0.040) | | (0.251) | | (0.039) |
| Age difference from cohort mean | 0.011 | Overweight | -0.137*** | Const. | -1.134*** |
| | (0.023) | | (0.036) | | (0.200) |
| Foreigner | -0.07 | Sports 1-2 /week | 0.041 | | |
| | (0.053) | | (0.043) | | |
| HH income | 0.771* | Sports 3-4 /week | 0.085 | | |
| | (0.398) | | (0.052) | | |
| Weekly earnings | 0.002 | Sports 5+ /week | 0.231*** | R-squared | 0.173 |
| | (0.025) | | (0.049) | N | 4847 |

Table 1 describes 'popular' teens by regressing our popularity measure on respondent attributes. The results show that popular teens tend to be white, from households with higher income, play sports, and they tend to be physically mature and attractive. In contrast, un-popular teens tend to be black, overweight, foreign, and new to the school.[4] The stereotypes found in these profiles support the validity of our network centrality as a measure of popularity.

## 3.3 GPA and Popularity

Table 2 present summary statistics of our key variables. GPA is measured in a 4-point scale with averages around 2.8 across subjects in both periods; %72 of our sample attended college while only %37 completed college. We will consider "good students" those with GPA above 3.0

[3]Note that about 15% of students were not properly matched to the nomination roster and thus their incoming friendship nominations were lost. A similar issue is that friendship nominations are truncated to a maximum of ten nominations. Costenbader and Valente (2003) find that, when working with a subsample of nodes, the eigenvector centrality is the most robust among the eight common centrality measures they consider: across several data sets, eigenvector centralities computed with 50% of the sample have an average correlations higher than 0.95 with eigenvector centralities computed with the full sample. This result gives us confidence on our centrality measure.

[4]Note that mathematics GPA does not predict popularity.

and as bad students those with GPA bellow 3.0. The bottom panel shows that the standardized popularity of good and bad students is essentially zero implying that both of these groups have average popularity. Appendix A1 presents summary statistics for all our variables; by design, this sample is representative of American high schools in 1995.

Table 2: Summary statistics

| | Mean | SD | N |
|---|---|---|---|
| GPA average 1995 | 2.85 | 0.79 | 6429 |
| Math GPA 1995 | 2.79 | 1.01 | 6148 |
| English GPA 1995 | 2.88 | 0.97 | 6139 |
| History GPA 1995 | 2.92 | 1.00 | 5532 |
| Science GPA 1995 | 2.89 | 1.00 | 5742 |
| GPA average 1996 | 2.82 | 0.75 | 7461 |
| Math GPA 1996 | 2.70 | 1.03 | 6757 |
| English GPA 1996 | 2.86 | 0.94 | 7352 |
| History GPA 1996 | 2.94 | 0.98 | 6417 |
| Science GPA 1996 | 2.87 | 0.98 | 6364 |
| Attended college | 0.72 | 0.45 | 6178 |
| Completed college | 0.37 | 0.48 | 6178 |
| ***Popularity of peers with:*** | | | |
| High GPA average ($>$ 3.0) | 0.08 | 1.14 | 31790 |
| Low GPA average ($<$ 3.0) | -0.07 | 0.83 | 29959 |
| High math GPA ($>$ 3.0) | 0.06 | 1.10 | 35533 |
| Low math GPA ($<$ 3.0) | -0.06 | 0.83 | 21998 |

## 3.4 Econometric Specification

Our identification strategy is based on variations in the popularity and GPA of students across adjacent cohorts within schools. Parents may select the school for their children, but they cannot select the cohort within the school. School fixed effects should therefore control for self-selection into schools as well as contextual variables at the school level (e.g., tutoring services). In addition, we include a dummy for parents who claim to have chosen their neighborhood in part for the school quality. We use lagged peer variables from 1995 (wave 1) to predict GPA in 1996 (wave 2) to control for reverse causality. We also include individuals' GPA in the first wave to control for individual ability, thus we essentially consider changes in GPA. Finally, we control for cohort mean GPA to account for the academic quality of the cohort.

Formally, our main specification will be a regression of the $GPA^{w_2}_{i,s,g}$ of individual $i$ in 1996 (wave 2) on the mean popularity of high-achievers, $\bar{P}^{w_1}_{s,g,Hgpa}$, and the mean popularity of low-achievers, $\bar{P}^{w_1}_{s,g,Lgpa}$, in school $s$ and cohort $g$ in 1995 (wave 1), controlling for the mean GPA in the cohort, $\overline{GPA}^{w_1}_{s,g}$:

$$GPA^{w_2}_{i,s,g} = c + z_s + \gamma_g + \beta X_i + \alpha_0 GPA^{w_1}_{i,s,g} + \alpha_1 \bar{P}^{w_1}_{s,g,y=0} + \alpha_2 \bar{P}^{w_1}_{s,g,y=1} + \alpha_3 \overline{GPA}^{w_1}_{s,g} \quad (1)$$

$X_i$ is a vector of demographics and household characteristics including gender, age (in months), race (black, hispanic, asian and other), a dummy for foreigners, household income and highest education of parents, a dummy for new students, and a dummy for parents who claim to have chosen their neighborhood in part for the school quality; $z_s$ is a school fixed effect and $\gamma_g$ is a cohort fixed effect. Our hypothesis is that $\alpha_2 > 0 > \alpha_1$, that is, individuals' GPA may increase with the popularity of peers with high GPA, and decrease with the popularity of peers with low GPA.

We also instrument the popularity of good and bad students based on peers' personal characteristics to provide additional support for the causal inferences based on school fixed effects and temporal lags. We need instruments that determine the popularity of peers but which do not influence the future GPA of others. Guided by the correlates of popularity found in Table 1, we use the percentage of new students, overweight, physically mature, and average sport participation among good students and, analogously, among bad students. These variables are only available for the InHome sample, thus we use the corresponding averages in the InHome sample to instrument for the popularity of peers in the InSchool sample (see Figure 1). The percentage of good and bad students who are physically mature should only affect their popularity but should not directly affect the GPA of others.

# 4 Results

## 4.1 Baseline results

Table 3: Regressions of total GPA (1996) on popularity of peers with high and low GPA

| | a | b | c | d | e |
|---|---|---|---|---|---|
| School and cohort FE | Yes | Yes | Yes | Yes | Yes |
| GPA 1995 | No | Yes | Yes | Yes | Yes |
| Demographics | No | Yes | Yes | Yes | Yes |
| Cohort average GPA | No | No | Yes | No | Yes |
| Friends average GPA | No | No | No | Yes | Yes |
| *Mean popularity among:* | | | | | |
| Peers with high GPA | 0.072*** | 0.057** | 0.061*** | 0.03 | 0.036 |
| | (0.026) | (0.024) | (0.024) | (0.024) | (0.024) |
| Peers with low GPA | -0.034 | -0.015 | -0.032 | 0.005 | -0.016 |
| | (0.032) | (0.030) | (0.030) | (0.030) | (0.030) |
| R-squared | 0.104 | 0.38 | 0.382 | 0.392 | 0.395 |
| N | 7256 | 6385 | 6385 | 5851 | 5851 |

Robust standard errors are in parentheses. High GPA is above 3.0; Low GPA is bellow 3.0. Peer variables are at the grade level.

**Total GPA.** Table 3 presents our baseline results for total GPA. In the first column we run a parsimonious regression using only school and cohort fixed effects along with the popularity of good and bad students; in subsequent columns we add controls sequentially to test the stability of our results. We find some evidence that the popularity of good students increases individuals' GPA the following year. Yet, this result is not robust to the inclusion of the mean GPA among direct friends (columns d and e).

**Subject GPA.** Looking more in depth, in Table 4 we estimate equation (1) separately for each academic subject using the popularity of good and bad students in each subject. We find that

Table 4: Subject GPA

| | English | Math | History | Science |
|---|---|---|---|---|
| *Mean popularity:* | | | | |
| Peers with high [subject] GPA | 0.014 | 0.215*** | 0.007 | 0.014 |
| | (0.037) | (0.045) | (0.043) | (0.044) |
| Peers with low [subject] GPA | -0.022 | -0.158*** | 0.011 | -0.05 |
| | (0.043) | (0.055) | (0.048) | (0.049) |
| R-squared | 0.242 | 0.213 | 0.257 | 0.213 |
| N | 5988 | 5523 | 4683 | 4894 |

Robust standard errors are in parentheses. High GPA is above 3.0; Low GPA is bellow 3.0. Peer variables are at the grade level.

all the effect of popularity on total GPA scores comes from mathematics. Indeed, a standard deviation increase in the popularity of good students increases individuals' math GPA next year by 0.2 GPA points. The same variation in the popularity of bad students decreases individuals' GPA the following year by 0.15 points.[5] For the other subjects, there is no effect from the popularity of good students. This may explain why results for total GPA are not fully robust in Table 3. One interpretation is that the logical nature of mathematics makes math GPA more sensitive to students' efforts compared to other subjects.

**Mathematics GPA.** In Table 5 we test the stability of the effects for mathematics. Results are stable to the inclusion of all the controls. In particular, we test the possibility that popular peers may not be more influential due to their social status, and that our results simply reflect that they reach more people. Thus, their influence may not be different from that of other peers but it may simply affect more friends. Hence, in columns d and e we control for the GPA of direct friends. The results are very similar suggesting that the effects we find do come from the "social status" of popular good/bad students at the cohort level *beyond* the number of individuals they reach directly.

[5]Note that Hoxby (2000) also finds much larger peer effects for mathematics than for reading.

Table 5: Regression Math GPA (1996) on popularity of peers with high and low GPA

| | a | b | c | d | e |
|---|---|---|---|---|---|
| School FE | Yes | Yes | Yes | Yes | Yes |
| Math GPA 1995 | No | Yes | Yes | Yes | Yes |
| Demographics | No | Yes | Yes | Yes | Yes |
| Cohort average math GPA | No | No | Yes | No | Yes |
| Friends average math GPA | No | No | No | Yes | Yes |
| *Mean popularity among:* | | | | | |
| Peers with high math GPA | 0.218*** | 0.206*** | 0.215*** | 0.178*** | 0.192*** |
| | (0.043) | (0.045) | (0.045) | (0.047) | (0.047) |
| Peers with low math GPA | -0.176*** | -0.143*** | -0.158*** | -0.103* | -0.124** |
| | (0.049) | (0.055) | (0.055) | (0.058) | (0.058) |
| R-squared | 0.087 | 0.211 | 0.213 | 0.214 | 0.217 |
| N | 6547 | 5523 | 5523 | 5046 | 5046 |

Robust standard errors are in parentheses. High GPA is above 3.0; Low GPA is bellow 3.0. Peer variables are at the grade level.

Table 6: Instrumental variables regression – GPA mathematics.

| | Math GPA 1996 |
|---|---|
| *Mean popularity among:* | |
| Peers with high math GPA | 0.537** |
| | (0.219) |
| Peers with low math GPA | -0.209 |
| | (0.190) |
| Sargan statistic | 5.769 |
| p-value Sargan | 0.45 |
| Hausman exogeneity test | 2.91 |
| P-value Haussman | 0.233 |
| R-Squared (centered) | 0.203 |
| N | 5510 |

Instrumental variables: physically mature; overweight, sports participation, and new students. IVs are computed both among good and bad students.

In Table 6 we present results from our instrumental variables estimation analogous to equation (1). In the bottom panel we can see that our instruments comfortably pass the validity test (p-value for Sargan statistic of 0.45). Also, the Hausman test cannot reject the exogeneity of our peers' popularity measures suggesting that a standard regression would be more appropriate. We still find the same results, albeit, we loose some statistical significance as would be expected. This supports a causal interpretation of our results.

## 4.2 Additional analyses

**Gender** We explore potential variations by gender. In Table 7 we estimate equation (1) separately for males and for females using peers popularity among males and females respectively. We find that the effect from the popularity of good students is somewhat stronger for males. Yet, the negative effect from the popularity of bad students seems to be driven by females! Several studies find that stereotypes about females' low achievement in mathematics can become self-full-filed prophecies (Steele and Ambady, 2006; Steffens et al., 2010). The popularity of girls with low math attainment may foster these patterns.

Table 7: Regression mathematics GPA by gender.

| | Females | Males |
|---|---|---|
| *Mean popularity:* | | |
| [gender] with high math GPA | 0.122** | 0.247*** |
| | (0.054) | (0.051) |
| [gender] with low math GPA | -0.170** | -0.017 |
| | (0.069) | (0.064) |

Robust standard errors are in parentheses. High GPA is above 3.0; Low GPA is bellow 3.0. Peer variables are at the grade level.

**College participation** We also explore whether our results extend to college participation. We run probit regression analogous to equation (1) for the probability of attending college and the probability of completing college respectively. In Table 8 we present the results. We find that the effect of the popularity of good/bad students in math extends to college attendance, and some evidence that it also affects college completion.

Table 8: College participation – probit average marginal effects.

| | Attended college | Completed college |
|---|---|---|
| *Mean popularity:* | | |
| Peers with high math GPA | 0.038** | 0.035* |
| | (0.018) | (0.020) |
| Peers with low math GPA | -0.037* | -0.017 |
| | (0.020) | (0.022) |

Robust standard errors are in parentheses. High GPA is above 3.0; Low GPA is bellow 3.0. Peer variables are at the grade level.

# 5 Conclusion

Adolescence is a tumultuous period in life during which individuals are supposed to focus on academics while they try to fit in with their peers. In this paper we find that the popularity (or social status) of good and bad students may considerably influence the academic performance of individuals in mathematics. The mean popularity of good students in mathematics increases the subsequent math GPA of other students, while the mean popularity of bad students has the opposite effect. This result is consistent with the hypothesis that popular students set social norms which influence the academic behavior of the cohort –at least in the case of mathematics. Note that mathematics are of particular importance because, compared to other subjects, math scores are better predictors of future income (Grogger and Eide, 1995; Weinberger, 2001). Indeed, further analysis suggests that our main results extend to college participation, that is the popularity of good students in mathematics increases the probability of individuals' attending and completing college and the opposite is true for the popularity of bad students.

These results are suggestive of the importance of social status in setting social trends. Barack and Michele Obama have already intuited this in their educational campaigns as they promote academics as 'cool'.

# References


Bifulco, R., J. M. Fletcher, and S. L. Ross (2011, February). The Effect of Classmate Characteristics on Post-Secondary Outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy 3*(1), 25–53.

Bursztyn, L. and R. Jensen (2015, July). How Does Peer Pressure Affect Educational Investments? *The Quarterly Journal of Economics 130*(3), 1329–1367.

Carrell, S. E. and R. L. Fullerton (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor* . . . .

Costenbader, E. and T. W. Valente (2003). The stability of centrality measures when networks are sampled. *Social Networks 25*(4), 283–307.

Fordham, S. and J. U. Ogbu (1986). Black students' school success: Coping with the "burden of 'acting white'". *The urban review 18*(3), 176–206.

Grogger, J. and E. Eide (1995). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources 30*(2), 280.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation.

Jackson, M. O. (2008). *Social and economic networks*. Princeton, NJ : Princeton University Press.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531.

McEwan, P. J. (2003, April). Peer effects on student achievement: evidence from Chile. *Economics of Education Review 22*(2), 131–141.

Steele, J. R. and N. Ambady (2006, July). "Math is Hard!" The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology 42*(4), 428–436.

Steffens, M. C., P. Jelenec, and P. Noack (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology 102*(4), 947–963.

Stinebrickner, R. and T. R. Stinebrickner (2006, September). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics 90*(8-9), 1435–1454.

Tyson, K. and W. Darity Jr (2005). It's not "a black thing": Understanding the burden of acting white and other dilemmas of high achievement. *American Sociological Review 70*(4), 582–605.

Weinberger, C. J. (2001). Is Teaching More Girls More Math the Key to Higher Wages? *Squaring Up: Policy Strategies to Raise ...*.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics 85*(1), 9–23.

# Appendix

Table A1: Summary Statistics

| | Mean | SD | N |
|---|---|---|---|
| GPA average 1995 | 2.85 | 0.79 | 6429 |
| Math GPA 1995 | 2.79 | 1.01 | 6148 |
| English GPA 1995 | 2.88 | 0.97 | 6139 |
| History GPA 1995 | 2.92 | 1.00 | 5532 |
| Science GPA 1995 | 2.89 | 1.00 | 5742 |
| GPA average 1996 | 2.82 | 0.75 | 7461 |
| Math GPA 1996 | 2.70 | 1.03 | 6757 |
| English GPA 1996 | 2.86 | 0.94 | 7352 |
| History GPA 1996 | 2.94 | 0.98 | 6417 |
| Science GPA 1996 | 2.87 | 0.98 | 6364 |
| Attended college | 0.72 | 0.45 | 6178 |
| Completed college | 0.37 | 0.48 | 6178 |
| Male | 0.48 | 0.50 | 7461 |
| Age | 16.16 | 1.59 | 7461 |
| Black | 0.21 | 0.41 | 7461 |
| Hispanic | 0.15 | 0.35 | 7461 |
| Asian | 0.05 | 0.22 | 7461 |
| Other | 0.01 | 0.11 | 7461 |
| Foreign | 0.06 | 0.25 | 7461 |
| New student in 1995 | 0.30 | 0.46 | 7461 |
| HH income ($1000) | 47.36 | 53.03 | 7461 |
| Move partly for school quality | 0.49 | 0.50 | 7461 |
| *Parents education* | | | |
| HS drop out | 0.14 | 0.35 | 7461 |
| High school | 0.29 | 0.45 | 7461 |
| Some college | 0.30 | 0.46 | 7461 |
| Graduated college | 0.16 | 0.37 | 7461 |
| Training beyond college | 0.10 | 0.30 | 7461 |

# Remittances, Local Income and Positional Concerns in Rural China

Alpaslan Akay[a,b,d], Olivier B. Bargain[c,d], Corrado Giulietti[d], Juan D. Robalino[e,d], Klaus F. Zimmermann[d,f]

[a] University of Gothenburg, Sweden

[b] LISER, Sweden

[c] Aix-Marseille University, CNRS & EHESS,

[d] IZA, Germany

[e] Cornell University, United States

[f] Bonn University, Germany

**Abstract**

This paper investigates the impact of remittances on the positional concerns of households in rural China. Using data from the survey of *Rural to Urban Migration in China* (RUMiC) we estimate a series of subjective well-being regressions to simultaneously explore positional concerns with respect to income and remittances. Our results show that the well-being of rural households decreases substantially with the average income of their reference group. However, their weill-being increases by a similar magnitude with the average remittances received by their reference group. These finding are robust to various model specifications, alternative definitions of reference groups, controls for the endogeneity of remittances and selective migration, as well as the use of migrants' net contribution to the household's income.

# 1 Introduction

For the past three decades China has been experiencing a massive movement of workers from rural towards urban areas. Recent estimates show that about 155 million people have left their rural residence to work in urban areas (Cai et al., 2011). Due to hukou restrictions, migrants' spouses and children are often left behind in villages, making remittances a crucial source of income to 'compensate' for the migrant's absence.[1] Estimates suggest that migrants sent about US $30 billion to rural areas in 2005 (Gong et al., 2008). Such large cash flows have important and complex effects not only on the welfare of family members left behind, but also on the development, income distribution and welfare of rural villages (e.g., Acosta et al., 2008; Howell, 2014). This paper investigates how remittances affect the positional concerns of rural households using subjective well-being (SWB hereafter) as a proxy for the experienced utility (Kahneman et al., 1997; Kahneman and Sugden, 2005).

An emerging strand of the literature has put forward the idea that individuals' SWB does not only depend on their absolute level of income but also on positional concerns, that is, on how individuals compare their income with that of other relevant people. The role of positional concerns is usually captured in SWB regressions by introducing the average income of the relevant reference group as an additional control variable. Evidence from developed countries suggests that the mean income of a reference group negatively affects individuals' well-being (e.g., in U.K: Clark and Oswald, 1996; Germany: Ferrer-i-Carbonell 2005; USA: Luttmer, 2005). These results suggest detrimental income comparisons between individuals and their reference groups. We will refer to such negative effects as a "status effect" (Clark et al., 2008).

However, empirical studies from transition countries have found some positive effects from the income of reference groups (e.g., in South Africa: Kingdon and Knight, 2007; Bookwalter

[1] Hukou is China's household registration system. Migrants from rural areas possess a rural hukou and they are typically unable to obtain an urban hukou. However, they are allowed to reside in a city as long as they are employed and can remain up to six months after unemployment.

and Dalenberg, 2010; Poland: Senik, 2005, 2008; Russia: Senik 2004; Ravallion and Lokshin, 2000). One possible explanation is the presence of altruistic feelings towards other members of the local community, or an increase in the supply of public goods. Another explanation is that the income of other people may act like a signal effect (or a "tunnel effect" as coined by Hirschman and Rothschild, 1973) for the individual's own income potential and prospects, thus, resulting in positive feelings (Senik, 2005). While we will not be able to disentangle the underlying mechanisms in detail, we will refer to such a positive effect as a "signal effect".

By investigating the effect of the remittances received by those in the reference group (we will use 'relative remittances' interchangeably) on individuals' SWB, we build upon and bring together the literature strands on positional concerns and remittances. The remittances literature focuses mainly on the absolute effect of remittances on well-being, particularly in relation to income inequality and poverty (e.g., Acosta et al., 2008; Akay et al., 2014). Remittances are expected to be positively associated with the well-being of individuals left behind since they represent an additional (or substitutive) source of income. However, remittances might change the income level of the reference group, thereby triggering a 'status' or a 'signal' effect.

Our analysis is based on the Rural to Urban Migration in China (RUMiC) dataset. By decomposing the overall household income into the part pertaining to remittances and the part pertaining to activities carried out in the village (henceforth: rural income), we find that rural households experience both status and signal effects at the same time. In particular, households exhibit a strong status effect with respect to rural income and an equally strong signal effect with respect to remittances. These two effects may be neglected when analyzing positional concerns using aggregate income measures. Results are robust to various definitions of the reference group. We also take into account self-selection into the pool of migrants by estimating selection equations and calculating the counterfactual income and expenditure distributions pertaining to migrants.

The paper is organized as follows: Section 2 provides information about the data and descriptive statistics. Section 3 outlines the econometric approach. Section 4 and 5 report the results from our baseline and additional models, respectively. Section 6 concludes.

# 2 Data

## 2.1 The RUMiC Dataset

We use the survey on Rural to Urban Migration in China (RUMiC), which consists of three distinct surveys: the Urban Household Survey, the Rural Household Survey, and the Migrant Household Survey.[2] Data were collected at the beginning of 2008, with most information (e.g., migration, income) referring to 2007. The Rural Household Survey covers the 9 largest migrant-sending provinces of China (Anhui, Chongqing, Guangdong, Hebei, Henan, Hubei, Jiangsu, Sichuan, and Zhejiang), and 82 counties as depicted in Figure 1.[3]

The dataset has rich information about demographic and socio-economic characteristics of household members, including questions on physical and mental health. We supplement the main dataset with a module of household income and expenditure collected in parallel to RUMiC, which contains information on income, remittances, assets, consumption and expenditure at the household level. Below we describe in detail the key variables used in the analysis.

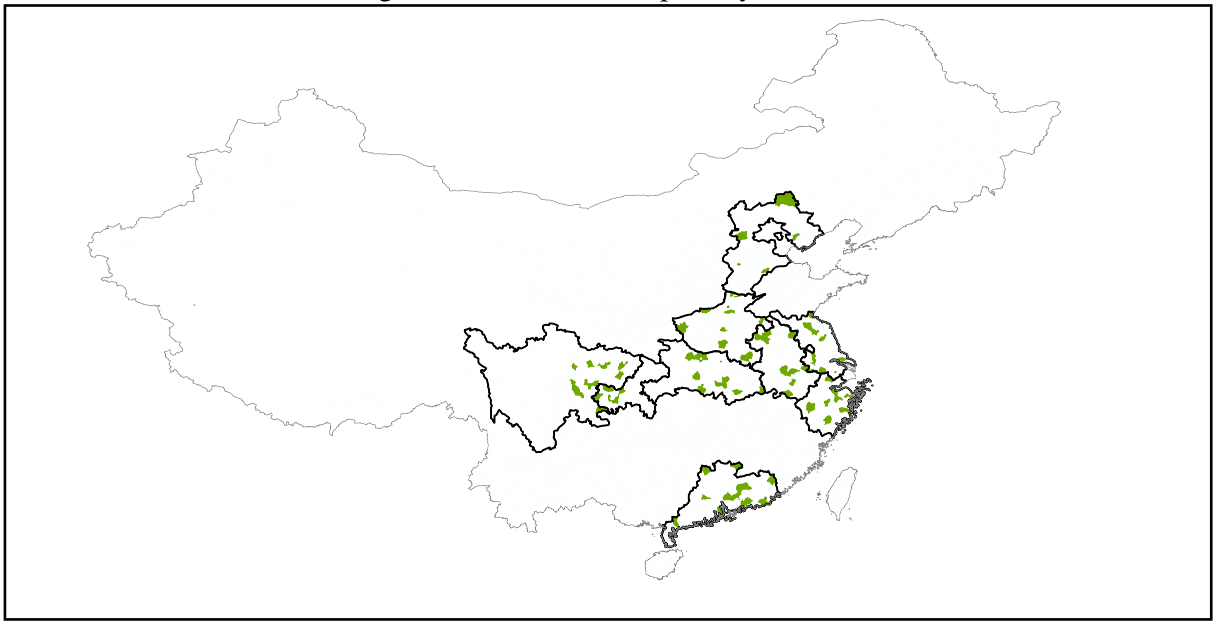
## 2.2 Measures of Well-being

The literature has identified several measures to proxy SWB (Frey and Stutzer, 2002; Dolan et al., 2008). These are generally based on survey questions about "happiness," "life-satisfaction" or "mental health." In order to measure SWB, we constructed an index based on the 12 mental health questions contained in the RUMiC's General Health Questionnaire (GHQ-12; see Ap-

[2] For an extensive description of RUMiC, see Akgüc et al. (2014).

[3] Counties are the administrative level above villages and below provinces

Figure 1: Counties sampled by RUMiC



Note: The green areas represent the counties sampled by RUMiC. The thick black borders delimit the nine provinces: Anhui, Chongqing, Guangdong, Hebei, Henan, Hubei, Jiangsu, Sichuan, and Zhejiang.

pendix A1). Information on GHQ-12 is available for all individuals aged 16 and above who were present at the time of the survey. The use of the GHQ-12 index has been widely accepted (see e.g., Clark and Oswald, 1994, 2002; Akay et al., 2012, 2013, 2014), and it is highly correlated with other SWB measures. Each question allows responses with scores between 0 (high levels) and 3 (low levels). To obtain the GHQ-12 index, we sum up the scores of the 12 questions, obtaining an index ranging from 0 to 36. We then reverse the scale so that 0 indicates the lowest SWB and 36 the highest level. One advantage of using an index that aggregates the 12 questions instead of using each question is that one has more variability in the SWB measure.

## 2.3 Sample and Descriptive Statistics

The sample consists of individuals aged between 16 and 70 who report SWB information. We cover a total of 11,624 individuals across 6,063 households.

**Migrants.** One of the key variables is individuals' migration status. This is constructed using the question, "How many months did you live away from the local township in 2007?" combined with an ancillary question on where the person lived during his or her absence ("If you lived outside the local township for more than three months in 2007, where did you live mainly?"). A migrant is defined as someone who lived in urban areas for at least six months during 2007. A household might have more than one migrant and the migrant can be the household head, spouse, or another household member. Given this definition, our data shows that 39% of households have at least one migrant. We also construct an indicator of whether individuals have migrated in the past (i.e., before 2007). Note that although migrants living in the city do not answer the SWB related questions, the survey respondent reports all socio-demographic and economic information for absent household members

**Rural Income and Remittances.** The other two key variables for this study are the households' overall income components: rural income and the remittances sent by migrants who live in the city. The rural income pertains to a mix of different activities conducted in the village including farming, wage employment, and self-employment. Income and remittances are measured at the household level in thousands of Chinese Yuans (CNY). While we explore different definitions, our preferred measure for rural income and remittances is the logarithm of per capita income using the modified OECD scale (the deflator is 1 for the household head, 0.5 for non-migrant adults and 0.3 for each child).

**Key statistics.** In Table 1 we report descriptive statistics of SWB, income and remittances (see Appendix A2 for the statistics of additional variables). We show statistics for the whole sample as well as separately for households who receive and do not receive remittances. The average SWB level is 28.08 (on our 0–36 scale). The difference between remittance receivers and non-receivers is very small, albeit statistically significant (28.27 versus 27.86 respectively, p-value=0.000). The mean level of remittances among receivers is 8,471 CNY, which constitutes more than a third of total income for these households. Average total income for these households is 24,990 CYN, compared to 26,785 CYN for households without remittances. However, the income difference between these two groups is not meaningful since household sizes differ systematically due to absent members who migrated. If we equalize incomes using the current household size, per capita incomes of the two groups become similar: 17,845 CYN for households receiving remittances versus 17,727 CYN for households without remittances.

Table 1: Summary statistics

| | Whole Sample | | | |
|---|---|---|---|---|
| | All | | HH Head | |
| | Mean | SD. | Mean | SD. |
| Subjective Well-Being (GHQ-12) | 28.081 | (5.18) | 28.648 | (4.89) |
| Household Income | 22.363 | (23.90) | 22.203 | (24.65) |
| Household per Capita Income | 14.962 | (15.43) | 15.351 | (16.52) |
| Remittances | 3.823 | (7.20) | 3.781 | (7.26) |
| Remittances per Capita | 3.365 | (5.22) | 3.357 | (5.25) |
| Observations | 11624 | | 6063 | |
| | Households with Remittances | | | |
| Subjective Well-Being (GHQ-12) | 27.856 | (5.25) | 28.412 | (5.01) |
| Household Income | 16.326 | (13.24) | 16.519 | (13.20) |
| Household per Capita Income | 11.628 | (9.91) | 12.003 | (10.27) |
| Remittances | 8.432 | (8.69) | 8.471 | (8.85) |
| Remittances per Capita | 6.217 | (6.73) | 6.282 | (6.80) |
| Observations | 5270 | | 2706 | |
| | Households without Remittances | | | |
| Subjective Well-Being (GHQ-12) | 28.267 | (5.11) | 28.838 | (4.79) |
| Household Income | 27.37 | (29.05) | 26.785 | (30.17) |
| Household per Capita Income | 17.727 | (18.36) | 18.05 | (19.79) |
| Observations | 6354 | | 3357 | |

Source: RUMiC 2008. Notes: Income and remittances are in 1,000 CNY/year.

# 3 Empirical Approach

## 3.1 Model Specification

Following the standard approach in the SWB literature, we identify positional concerns by regressing SWB on the average income of the 'reference group' (e.g., individuals in the same county) while controlling for individuals' absolute income level and detailed socio-economic characteristics (e.g., Ferrer-i-Carbonell, 2005; Akay et al., 2012, 2013). We decompose this relative income in two components: the relative rural income and the relative remittances. Hence, our specification includes measures of individuals' local income $Y_i$ and remittances $R_i$ as well as the reference group's average local income $\bar{Y}_j$ and average remittances $\bar{R}_j$. Formally we have:

$$SWB_{ij} = \alpha_1 Y_i + \alpha_2 R_i + \rho_1 \bar{Y}_j + \rho_2 \bar{R}_j + X_i \beta + \eta_k + \epsilon_{ij} \tag{1}$$

where $X_i$, captures demographic characteristics such as age, sex, marital status, and health status; and the term $\eta_k$ refers to indicators for provinces. The standard errors are clustered at the household level. The key parameters in our analysis are $\rho_1$ and $\rho_2$, which capture respectively the effects of relative rural income and relative remittance. The expected signs of $\rho_1$ and $\rho_2$ are not known a priori. Rural income or remittances of the reference group could correlate negatively with SWB possibly due to income comparisons. As described in the introduction, we refer to this situation as the "status effect" with respect to a particular source of income. On the contrary, a positive coefficient would be denoted as a "signal effect". The model presented in Equation (1) might be subject to selection bias. Indeed, one issue could be that preferences towards status might influence the incentives to migrate. We investigate this in Section 5.

## 3.2 Defining Reference Groups

One crucial issue is how to define the reference group. While reference groups are usually unknown, the literature suggests two distinct approaches to identify groups to which individuals refer. The first method involves using socio-demographic similarities (e.g., Clark and Oswald, 1996; McBride, 2001; Ferrer-i-Carbonell, 2005; Luttmer, 2005). The second method directly asks people to whom they compare themselves (Clark and Senik, 2010). In a survey among rural households in China, Knight et al. (2009) find that 70% of the respondents report that they compare their income with that of village members. In our work we also use as the reference the population of the region where individuals live. Due to sample size limitations we cannot use village residents as the reference group (on average, only 10 households are sampled in each village). Instead, we refer to a slightly larger orbit of comparison by using "counties" as defined in the data section. We also explore the sensitivity of our results to narrower reference groups by taking into account other population characteristics (e.g., wage workers within the same county).

# 4 Results

As a preliminary step, we analyze the general determinants of SWB in rural China. Then, we outline results for positional concerns with respect to rural income and remittances, we check the sensitivity of the results to alternative reference groups, and we investigate the role of income inequality. Finally we explore observed heterogeneity.

## 4.1 Determinants of Subjective Well-Being in Rural China

In Appendix (A3) we present the results of a standard OLS regression of SWB. The scope of this regression is to show how our estimates compare with those of existing SWB studies. We present results for the whole sample as well as separately for households receiving and not receiving remittances. The signs, magnitudes, and statistical significance of the estimates pertaining to socio-economic and demographic characteristics align with other studies (e.g., Dolan et al., 2008; Knight et al., 2009; Akay et al., 2012). Results are also similar across households who receive remittances and those who do not, albeit there are differences in the magnitude and significance of some parameter estimates.

**Rural Income and Remittances.** In the first panel of Table 2, we include the two components of household income (rural income and remittances), indicators for economic activity, and other wealth measures. As one would expect, rural income is positively correlated with SWB. The magnitude of this correlation is similar across specifications and is usually statistically significant, with the exception of the specification related to household members in the remittance-receiving group. Such positive correlation is consistent with the results for developed countries (e.g., Dolan et al., 2008) and also with previous evidence from China (Knight et al., 2009; Knight and Gunatilaka, 2010; Akay et al., 2012). The effect from remittances is close to that of rural income and, in fact, the difference between the two estimates is not statistically significant (0.210 versus 0.175, p-value=0.7 for the whole sample). Our regressions include additional economic and wealth-related variables. For example, being a wage worker, a farmer, or self-employed is associated with higher well-being than the reference category (which is the group formed by the inactive population and those who do household work). However, such effects are most important for households that are not receiving remittances. Also, working more hours leads to lower well-being, particularly for households without remittances (see Pouwels et al., 2008). Conditional on other wealth and income measures, land size is negatively correlated with SWB among remittance-receiving households. One possible interpretation is that individuals in this group need additional labor in order to maintain the land and carry out agricultural activities after the migrant has left. Finally, and perhaps as expected, both the size and the value of the house are positively correlated with SWB, with a stronger effect for house size.

**Socio-demographic Characteristics.** The estimates for age and age-squared suggest the existence of a U-shaped relationship, which very much aligns with international evidence (see Blanchflower and Oswald, 2008). Since the sample of left-behind individuals is relatively old (the average age is 48), the minimum point of the age effect is located between 55 and 60. Males report higher SWB compared to females, although it is important to point out that the group of males staying in rural regions is strongly self-selected. In line with other studies, we observe a strong positive effect from marriage (e.g., Helliwell, 2003). Health is a strong predictor of SWB (the omitted category is "very good health"). Weight appears to be positively correlated with SWB, while height seems to be unrelated to it. Education is another strong predictor of SWB, once again in line with other studies (e.g., Helliwell, 2003). The few households holding urban hukou (6% of the total) experience higher SWB, though the effect is found only among those who receive remittances. While household size is strongly associated with SWB, we found only a weak relationship with the number of children, in line with the findings reported in Dolan et al. (2008).

**Migration Characteristics.** We control for a rich set of variables capturing the households' migration characteristics. First, we include dummy variables to capture the number of migrants in the households (the base category is "no migrants"). In general, having one or more migrants in the household is negatively associated with SWB. Recent literature has shown that the absence of household members due to migration can have both positive and negative effects on well-being determinants, such as education and self-employment (Biavaschi et al. 2015; Giulietti et al. 2013). Since these factors are already controlled for in the regression, we interpret the negative estimate as the psychological cost associated with the migrant's absence. A back of the envelope calculation suggest that, on average, a migrant's absence has a larger cost in terms of well-being than the resulting benefits from remittance.

## 4.2 Positional concerns in Rural China

**Baseline Results.** In Table 2 we present the results from the full model presented in Equation (1). We are mainly interested in the effect of relative income and relative remittances. The

Table 2: SWB equation

| *Dependent variable: GHQ-12* | Whole Sample | Only HH Heads |
|---|---|---|
| Log HH income pc | 0.340*** | 0.375*** |
| | (0.089) | (0.097) |
| Log remittances pc | 0.07 | 0.067 |
| | (0.072) | (0.079) |
| County mean Log HH income pc | -0.840*** | -0.709*** |
| | (0.219) | (0.238) |
| County mean Log remittances pc | 0.873*** | 0.902*** |
| | (0.186) | (0.200) |
| R-squared | 0.238 | 0.196 |
| Observations | 11624 | 6063 |

Notes: Robust standard errors clustered at the household level are shown in parentheses. */**/*** indicate significance at the 10%, 5%, and 1% levels. Regressions include all covariates from Table A3. Income is measured in 1000 CNY. Per capita income is calculated using the modified OECD equivalence scale.

average rural income of the reference group is statistically significant, with a large negative magnitude (-0.840). Hence, results point to a status effect experienced by rural household members. We also find a large positive effect from relative remittances (0.873). As mentioned in the introduction, this positive effect could be due to an improvement in public goods or even altruism. Yet, the fact the we find a negative effect with respect to local income (which could have also been subject to altruism or improved public goods) suggests that this is a "signaling effect" of the potential opportunities of migrations, possibly for both potential migrants and potential remittance receivers.

In sum, our baseline results in Table 2 imply that rural individuals experience a substantial negative effect with respect to relative income and an equally large positive effect with respect to relative remittances. The result is very similar when we narrow the sample to include only household heads who may feel more responsible for the household's finances.

**Reference Groups.** In Table 3 we explore the sensitivity of relative remittances to the definition of the reference group. To do so, we keep the county as the regional orbit of comparison and add some additional criteria. The first two alternative reference groups involve similarity in age and wage employment status (columns one and two) and produce very similar results to

Table 3: SWB equations with alternative reference groups

| *Dependent variable: GHQ-12* | Reference Groups | | |
|---|---|---|---|
| | County + Wage Workers | County + Age<40 | County + HH with Migrants |
| Log HH income pc | 0.353*** | 0.331*** | 0.285*** |
| | (0.090) | (0.089) | (0.090) |
| Log remittances pc | 0.073 | 0.052 | -0.011 |
| | (0.072) | (0.072) | (0.074) |
| Mean Log HH income pc in reference group | -0.884*** | -0.755*** | -0.423** |
| | (0.212) | (0.214) | (0.210) |
| Mean Log remittances pc in reference group | 0.810*** | 0.921*** | 1.097*** |
| | (0.183) | (0.169) | (0.181) |
| R-squared | 0.238 | 0.239 | 0.239 |
| Observations | 11624 | 11624 | 11624 |

Notes: Robust standard errors clustered at the household level are shown in parentheses. */**/*** indicate significance at the 10%, 5%, and 1% levels. Regressions include all covariates from Table A3. Income is measured in 1000 CNY. Per capita income is calculated using the modified OECD equivalence scale.

our benchmark. In the third column we calculate relative income and remittances among households with and without migrants. The results are slightly different to those in the first column, but the same patterns hold.

**Does income inequality explain the results?** Due to their substantial size, remittances are expected to affect the income distribution within the receiving regions. The empirical evidence, mostly based on international migration, suggests that the relationship between remittances and income inequality is unclear. Moreover, income inequality might have an effect on SWB and this could be confounded with the 'signal effect' (on this point see, e.g., Senik, 2008). To address this question, we first calculate the Gini index in the county using both the pre- and post-remittances per capita household income. We find that the remittances' contribution to income inequality varies across regions: around 46% of the counties exhibit higher income inequality after receiving remittances. Results in Table 4 show that accounting for the change in income inequality before and after remittances does not affect our baseline results. The sign of the correlation with the change in inequality is negative –suggesting that increases in inequality due to remittances reduce SWB– but is statistically insignificant.

Table 4: SWB and change in inequality

| | GHQ-12 |
|---|---|
| Log HH income pc | 0.342*** |
| | (0.089) |
| Log remittances pc | 0.067 |
| | (0.072) |
| County mean Log HH income pc | -0.837*** |
| | (0.220) |
| County mean Log remittances pc | 0.804*** |
| | (0.198) |
| Change in income inequality due to remittances | -1.728 |
| | (1.42) |
| R-squared | 0.239 |
| Observations | 11624 |

Notes: Robust standard errors clustered at the household level are shown in parentheses. */**/*** indicate significance at the 10%, 5%, and 1% levels. Includes all covariates from Table A3. Income is measured in 1000 CNY. Per capita income is calculated using the modified OECD equivalence scale.

## 4.3 Heterogeneity

Our sample comprises heterogeneous individuals and regions. We therefore investigate how our main results differ across various groups through a simple modification of the baseline empirical model. We first define a dummy variable indicating a specific individual or county characteristic, $D_i$, and then we interact $D_i$ and $(1 - D_i)$ with the absolute and relative rural income and remittance variables (i.e., interaction terms are arranged to show the total effect for each group as oppose to the difference in effects). Tables 5 presents the estimates of the interactions.

In the first panel of Table (5) we find a much stronger status effect from relative income among the young. This may suggest that younger people are more competitive. We also find a much stronger signal effect from relative remittances among the older. In the context of rural China, remittances are often sent to support elderly parents, thus, the signal effect may be driven by the potential benefits that the elder can get from remittances. In the second panel we find that counties with a higher percentage of wage workers seem to have stronger status effects from local income, but similar signal effects from remittances. In the third panel we find that in counties with income above the median of all counties the status effect is much

Table 5: Heterogeneity

| *Dependent variable: GHQ-12* | Older than 40 Years Old | | % of Wageworkers in County above Median | | Relatively Rich County | | Remittances Decreased County Inequality | |
|---|---|---|---|---|---|---|---|---|
| | D=1 | D=0 | D=1 | D=0 | D=1 | D=0 | D=1 | D=0 |
| Log HH income pc | 0.419*** | 0.147 | 0.534*** | 0.152 | 0.442*** | 0.242** | 0.299*** | 0.389*** |
| | (0.099) | (0.144) | (0.117) | (0.122) | (0.117) | (0.122) | (0.115) | (0.123) |
| Log remittances pc | 0.044 | 0.1 | 0.101 | 0.059 | 0.074 | 0.072 | 0.05 | 0.09 |
| | (0.077) | (0.141) | (0.095) | (0.094) | (0.098) | (0.091) | (0.092) | (0.098) |
| County mean | -0.547** | -1.568*** | -1.410*** | -0.717** | -1.511*** | -0.092 | -0.616** | -1.090*** |
| log HH income pc | (0.239) | (0.339) | (0.366) | (0.343) | (0.461) | (0.405) | (0.285) | (0.308) |
| County mean | 1.101*** | 0.185 | 0.662** | 0.975*** | 0.695** | 0.883*** | 0.872*** | 0.812*** |
| log remittances pc | (0.194) | (0.370) | (0.326) | (0.235) | (0.282) | (0.286) | (0.251) | (0.293) |

Notes: Robust standard errors clustered at the household level are shown in parentheses. */**/*** indicate significance at the 10%, 5%, and 1% levels. Regressions include all covariates from Table A3. Income is measured in 1000 CNY. Per capita income is calculated using the modified OECD equivalence scale.

stronger, and in fact it is close to zero for relatively poorer counties. It seems that a higher economic level in the region relates to higher positional concerns (Clark et al., 2008). Yet, the signal effect from remittances is similar. In the fourth panel we investigate the role of income inequality by interacting an indicator for counties where remittance flows reduced the inequality from rural incomes. We find a lower status effect with respect to income in regions where income inequality has decreased. Yet, the effects from relative remittances are very similar to the baseline model, implying once more that income inequality does not play a role in the signal effect.

# 5 Robustness to Self-Selection and Net Remittances

In this section we explore the potential confounding role of selectivity into migration. Then, we present an analysis accounting for the migrant's counterfactual rural income and expenditures.

## 5.1 Selective Migration

The decision to migrate may induce a sample selection bias as migrants, and households with migrants, may be intrinsically different. For instance, people's intrinsic preferences towards status may generate a strong incentive to migrate (Stark and Yitzhaki, 1988; Stark, 2006). One of the features of our sample is that we have socio-demographic characteristics (except SWB information) for individuals who are currently migrants in the urban areas. This allows us to specify a selection model as follows. First we estimate a probit model for the probability of migrating:

$$s_i = 1(Z_i\gamma + w_i > 0) \tag{2}$$

where $s_i$ is an indicator taking the value 1 if the individual is not a migrant (which in our data corresponds to whether the SWB of individual $i$ is observed). Then we estimate a SWB equation correcting for the decision to migrate with an inverse mills ratio to account for the selection bias:

$$E(SWB_i|s_i = 1) = \alpha_1 Y_i + \alpha_2 R_i + \rho_1 \bar{Y} + \rho_2 \bar{R} + X_i\beta + \eta_k + \rho_{w\epsilon}\sigma \frac{\phi(Z_i\gamma)}{1 - \Phi(Z_i\gamma)} \tag{3}$$

where $\rho_{w\epsilon}$ is the correlation between the selection equation and the SWB equation. Identification requires at least one exclusion restriction which will be used in the selection equation and excluded from the target SWB equation, i.e., variables that affect migration but not SWB. We use the following three instruments: (i) distance to the nearest port or station, (ii) distance to the nearest government building, and (iii) order of bird. Individuals are more likely to migrate if stations or government buildings are closer, and first-borns are less likely to migrate since they are often responsible for taking care of the family. Hence, instruments are thought to influence migration without having a direct effect on SWB.

The correlation between the equations is significant (0.108). Yet, the estimates presented in Table 6 are very similar to our baseline estimates suggesting that our main results are robust to self-selection into migration.

Table 6: Sample selection correction model

| | GHQ-12 |
|---|---|
| Log HH income pc | 0.197** |
| | (0.084) |
| Log remittances pc | -0.079 |
| | (0.066) |
| County mean Log HH income pc | -0.741*** |
| | (0.217) |
| County mean Log remittances pc | 0.822*** |
| | (0.186) |
| Rho (correlation between equations) | 0.108* |
| | (0.062) |
| Observations | 16156 |

Notes: Robust standard errors clustered at the household level are shown in parentheses. */**/*** indicate significance at the 10%, 5%, and 1% levels. Includes all covariates from Table A3. Income is measured in 1000 CNY. Per capita income is calculated using the modified OECD equivalence scale.

## 5.2 Counterfactual Income and Net Remittances

Migrants' contribution to the household through remittances does not take into account the counterfactual rural income he would have contributed had he not migrated. The signaling effect could even be altruism towards the families who receive a compensation for migrants' prior rural income . Failure to account for the potential income contribution of absent migrants might lead to misinterpretations of the absolute or relative remittances. In this section, we simulate migrants' counterfactual income and expenditures in order to calculate net remittances, that is, migrants' net contribution to the household income (Barham and Boucher, 1998; Howell, 2014). Similar to the previous exercise, we first model the decision to migrate:

$$s_i = 1(g_i\psi + \mu > 0) \tag{4}$$

Then we estimate counterfactual income correcting for the migration decision with an inverse mills ratio:

$$E(Y_i|s_i = 0) = X_i\beta + \rho_{\mu\epsilon}\sigma\frac{-\phi(g_i\psi)}{\Phi(g_i\psi)} \tag{5}$$

where $\rho_{\mu\epsilon}$ is the correlation between equations. The predicted value from Equation (5) provides an estimate of what the migrant's income would have been had he or she not migrated. We use an identical procedure to obtain the counterfactual expenditures of migrants. In modeling migration we use the same instruments as above, i.e., distances and birth order. To obtain the net remittances, we subtract the difference between the counterfactual income and counterfactual expenditure from the actual remittances. A significant portion of net remittances are negative since some migrants contribute more to the household income in the counterfactual situation, i.e., had they not migrated.

As a last step, we calculate the mean net remittances within the reference group and use this to estimate the baseline model specification in Equation (1). Table 7 reports the results. Since there are several negative values in the net remittances, we use values in level and not in log, hence magnitudes are not directly comparable with the previous estimates. However, the pattern of the estimates is in line with the baseline: we find a negative effect from relative income and a positive effect from relative remittances.

Table 7: SWB equation with net remittances

| | GHQ-12 |
|---|---|
| HH income pc | 0.015*** |
| | (0.004) |
| Net remittances pc | 0.012 |
| | (0.008) |
| County mean HH income pc | -0.037** |
| | (0.016) |
| County mean net remittances pc | 0.031** |
| | (0.015) |
| R-Squared | 0.235 |
| Observations | 11542 |

Notes: Robust standard errors clustered at the household level are shown in parentheses. */**/*** indicate significance at the 10%, 5%, and 1% levels. Regressions include all covariates from Table A3. Income is measured in 1000 CNY. Per capita income is calculated using the modified OECD equivalence scale.

# 6 Conclusions

To the best of our knowledge, this is the first paper investigating the effect of remittances on households' positional concerns. Using the survey of Rural to Urban Migration in China, we simultaneously investigate the role of relative income and relative remittances on individuals' SWB in rural China. Our results suggest that the relative rural income has a strong negative effect on SWB while the relative remittances have a positive effect of similar magnitude. While local income seems to induce detrimental income comparisons, the remittances received by peers seem to signal opportunities and produce positive feelings (although we cannot fully disentangle alternative explanations such as altruism or improved public goods). These results are robust to several definitions of reference groups, changes in income inequality due to remittances, as well as to sample-selection bias arising from the decision to migrate.

The main implication of our paper is that the role of relative income on individuals' SWB depends on the source and context of income, in this case the part accruing to remittances and the part generated by rural activities. Failing to disentangle these aspects would confound the status and signal effects potentially experienced by rural households.

# References


Acosta, P., C. Calderon, P. Fanjzylber, and H. Lopez (2008). What is the Impact of International Remittances on Poverty and Inequality in Latin America? *World Development* 36(1): 89-114.

Akay, A., O. Bargain, and K.F. Zimmermann (2012). Relative concerns of Rural-to-Urban Migrants in China, *Journal of Economic Behavior and Organization* 81: 421-441.

Akay, A., C. Giulietti, J.D. Robalino, and K.F. Zimmermann (2014). Remittances and well-being among rural-to-urban migrants in China, *Review of Economics of the Household* 12(3): 517-546.

Akay, A., G. Karabulut, and P. Martinsson (2013). The Effect of Religiosity and Religious Festivals on Positional Concerns: An Experimental Investigation of Ramadan. *Applied Economics* 45(27): 3914-3921.

Akgüc, M., C. Giulietti, and K.F. Zimmermann (2014). The RUMiC Longitudinal Survey: Fostering Research on Labor Markets in China, *IZA Journal of Labor and Development* 3: 5.

Barham, B., and S. Boucher (1998). Migration, remittances and inequality: estimating the net effects of migration on income distribution, *Journal of Development Economics* 55: 307-331.

Biavaschi, C., C. Giulietti, and K.F. Zimmermann (2015). Sibling Influence on the Human Capital of the Left Behind, *Journal of Human Capital* forthcoming.

Blanchflower, D.G., and A.J. Oswald (2008). Is Well-being U-Shaped over the Life Cycle?, *Social Science & Medicine* 66: 1733-1749.

Bookwalter, J., and D.R. Dalenberg (2010). Relative to What or Whom? The Importance of Norms and Relative Standing, *World Development* 38: 345-355.

Cai, F., Y. Du, and D. Wang (Eds.) (2011). *Zhongguo renkou yu laodong wenti baogao No.12 (Report on China's Population and Labour)*, Beijing: Shehui kexue wenxian chubanshe.

Clark, A., and A. Oswald (1994). Unhappiness and unemployment, *Economic Journal* 104: 648-659.

Clark, A.E. and A.J. Oswald (1996). Satisfaction and comparison income, *Journal of Public Economics* 61: 359-381.

Clark, A., and A. Oswald (2002). A simple statistical method for measuring how life events affect happiness, *International Journal of Epidemiology* 31(6): 1139-1144.

Clark, A.E., and C. Senik (2010). Who Compares to Whom? The Anatomy of Income Comparisons in Europe, *Economic Journal* 120: 573-594.

Clark, A.E., P. Frijters, and M.A. Shields (2008). Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles, *Journal of Economic Literature* 46: 95-144.

Dolan, P., T. Peasgood, and M. White (2008). Do We Really Know What Makes Us Happy? A Review of the Economic Literature on the Factors Associated with SWB, *Journal of Economic Psychology* 29: 94-122.

Ferrer-i-Carbonell, A. (2005). Income and well-being: An empirical analysis of the comparison income effect, *Journal of Public Economics* 89: 997-1019.

Frey, B., and A. Stutzer (2002). What can economists learn from happiness research? *Journal of Economic literature* 40(2): 402-435.

Giulietti, C, J. Wahba, and K.F. Zimmermann (2013). Entrepreneurship of the Left Behind, *Research in Labor Economics* 37: 65-92.

Goldberg, D. (1978). *Manual of the General Health Questionnaire.* Windsor, England: NFER Publishing; 1978.

Gong, X., S. Kong, S. Li, and X. Meng (2008). Rural-urban migrants: A driving force for growth, China's Dilemma: Economic Growth. In Song, L., and Thye Woo, W. (Eds.) *The Environment and Climate Change* (pp. 110-152). Canberra: Asia Pacific Press.

Helliwell, J. (2003). How's life? Combining individual and national variables to explain subjective well-being, *Economic Modelling* 20(2): 331-360.

Hirschman, A., and M. Rothschild (1973): The changing tolerance for income inequality in the course of economic development, *Quarterly Journal of Economics* 87: 544-565.

Howell, A. (2014). Ethnic Migration, Remittances and Rural Development in China's Least Developed Areas: A Case of "Immiserizing Growth"?, *Working Paper.*

Kahneman, D., and R. Sugden (2005). Experienced utility as a standard of policy evaluation, *Environmental and Resource Economics* 32(1): 161-181.

Kahneman, D., P. Wakker, and R. Sarin (1997). Back to Bentham? Explorations of Experienced Utility, *Quarterly Journal of Economics*, 112(2): 375-405.

Kingdon, G., and J. Knight (2007). Community, comparisons and subjective well-being in a divided society, *Journal of Economic Behavior and Organization* 64: 69-90.

Knight J., L. Song, and R. Gunatilaka (2009). Subjective well-being and its determinants in rural China, *China Economic Review* 20: 635-649.

Knight, J. and R. Gunatilaka (2010). Great expectations? The Subjective Well-being of Rural-Urban Migrants in China, *World Development* 38: 113-124.

Luttmer, E. (2005). Neighbors as negatives: Relative earnings and well-being, *Quarterly Journal of Economics* 120: 963-1002.

McBride, M. (2001). Relative-income effects on subjective well-being in the cross-section, *Journal of Economic Behavior and Organization* 45: 251-278.

Pouwels, B., J. Siegers, and J.D. Vlasblom (2008). Income, working hours, and happiness, *Economics Letters* 99(1): 72-74.

Ravallion, M., and M. Lokshin (2000). Who Wants to Redistribute? The Tunnel-Effect in 1990s Russia, *Journal of Public Economics* 76: 87-104.

Senik, C. (2004). When Information Dominates Comparison: Learning from Russian subjective panel data, *Journal of Public Economics* 88: 2099-2123.

Senik, C. (2005). Income distribution and well-being: what can we learn from subjective data?, *Journal of Economic Surveys* 19: 43-63.

Senik, C. (2008). Ambition and Jealousy. Income Interactions in the ”Old Europe” versus the ”New Europe” and the United States, *Economica* 75: 495-513.

Stark, O. (2006). Inequality and migration: A behavioral link, *Economics Letters* 91: 146-152.

Stark, O., and S. Yitzhaki (1988). Labour migration as a response to relative deprivation, Journal of Population Economics 1(1): 57-70.

Wu, Z. (2008). Relative Income Positions and Labor Migration: A Panel Study Based on a Rural Household Survey in China, *Osaka Discussion Papers in Economics and Business* 8-24.

# Appendix

Table A1: General Health Questionnaire (GHQ-12)

| | |
|---|---|
| 1- When you are doing something, do you find that<br>(1) Can concentrate; (2) Attention occasionally diverted; (3) Attention sometimes diverted; (4) Attention frequently diverted, cannot concentrate | 7- Are you able to enjoy day-to-day activities?<br>(1) Very interesting; (2) Fairly interesting;<br>(3) Not very interesting; (4) Not interesting at all |
| 2- Do you often lose sleep over worry?<br>(1) Not at all; (2) Occasionally; (3) Fairly often; (4) Very often | 8- Are you able to face problems?<br>(1) Never; (2) Seldom; (3) Sometimes; (4) Always |
| 3 - Can you play a useful part in things?<br>(1) Always can; (2) Can play some positive roles; (3) Can play positive roles poorly; (4) Cannot play a positive role | 9- Do you feel depressed?<br>(1) Not at all; (2) A little bit; (3) Fairly seriously;<br>(4) Very seriously |
| 4- Are you capable of making decisions?<br>(1) Always have own opinions; (2) Sometimes have own opinions; (3) Do not have many own opinions; 4) Do not have any personal opinion at all | 10- Do you always lack confidence?<br>(1) Not at all; (2) A little bit; (3) Fairly seriously;<br>(4) Very seriously |
| 5- Are you constantly under strain?<br>(1) Never; (2) Sometimes; (3) Fairly often; (4) Very often | 11- Do you often think that you have no value?<br>(1) Not at all; (2) A little bit;<br>(3) Fairly seriously; (4) Very seriously |
| 6- Do you feel you couldn't overcome difficulties?<br>(1) Never; (2) Sometimes; (3) Fairly often; (4) Very often | 12- Are you happy when you consider each aspect of your life?<br>(1) Very happy; (2) Fairly happy;<br>(3) Not very happy; (4) Not happy at all |

See Goldberg (1978) for questions and use of the GHQ.

## Table A2: Descriptive Statistics of Control Variables

| | Whole Sample | | | | Households without Remittances | | | | Households with Remittances | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All | | HH Head | | All | | HH Head | | All | | HH Head | |
| | Mean | SD. | Mean | SD. | Mean | SD. | Mean | SD. | Mean | SD. | Mean | SD. |
| Subjective Well-Being (GHQ-12) | 28.081 | (5.179) | 28.648 | (4.891) | 28.267 | (5.112) | 28.838 | (4.789) | 27.856 | (5.251) | 28.412 | (5.006) |
| Age | 48.106 | (11.029) | 50.874 | (9.235) | 47.494 | (11.467) | 50.301 | (9.807) | 48.845 | (10.431) | 51.585 | (8.421) |
| Male | 0.557 | (0.497) | 0.954 | (0.209) | 0.573 | (0.495) | 0.961 | (0.193) | 0.538 | (0.499) | 0.945 | (0.227) |
| Married | 0.934 | (0.249) | 0.954 | (0.210) | 0.924 | (0.265) | 0.952 | (0.214) | 0.946 | (0.227) | 0.956 | (0.205) |
| Good Health | 0.491 | (0.500) | 0.497 | (0.500) | 0.495 | (0.500) | 0.499 | (0.500) | 0.487 | (0.500) | 0.494 | (0.500) |
| Average Health | 0.232 | (0.422) | 0.231 | (0.421) | 0.22 | (0.415) | 0.214 | (0.410) | 0.247 | (0.431) | 0.251 | (0.434) |
| Poor Health | 0.053 | (0.225) | 0.045 | (0.208) | 0.05 | (0.218) | 0.045 | (0.207) | 0.057 | (0.232) | 0.046 | (0.210) |
| Years of Education | 6.742 | (3.065) | 7.279 | (2.629) | 6.964 | (3.122) | 7.442 | (2.698) | 6.476 | (2.974) | 7.077 | (2.527) |
| Urban Hukou | 0.06 | (0.238) | 0.056 | (0.230) | 0.082 | (0.274) | 0.077 | (0.267) | 0.035 | (0.183) | 0.03 | (0.170) |
| One Child | 0.262 | (0.440) | 0.247 | (0.431) | 0.302 | (0.459) | 0.291 | (0.454) | 0.215 | (0.411) | 0.192 | (0.394) |
| Two Children | 0.402 | (0.490) | 0.423 | (0.494) | 0.392 | (0.488) | 0.421 | (0.494) | 0.415 | (0.493) | 0.426 | (0.495) |
| More than Two Children | 0.28 | (0.449) | 0.303 | (0.459) | 0.243 | (0.429) | 0.262 | (0.440) | 0.326 | (0.469) | 0.353 | (0.478) |
| Number HH Members | 4.06 | (1.376) | 3.965 | (1.393) | 3.749 | (1.306) | 3.629 | (1.293) | 4.436 | (1.363) | 4.381 | (1.401) |
| Relationship to Head: Head | 0.522 | (0.500) | 1 | 0.000 | 0.528 | (0.499) | 1 | 0.000 | 0.513 | (0.500) | 1 | 0.000 |
| Relationship to Head: Spouse | 0.374 | (0.484) | 0 | 0.000 | 0.362 | (0.481) | 0 | 0.000 | 0.389 | (0.488) | 0 | 0.000 |
| Relationship to Head: Child | 0.083 | (0.276) | 0 | 0.000 | 0.094 | (0.292) | 0 | 0.000 | 0.07 | (0.255) | 0 | 0.000 |
| Spouse is a Migrant | 0.018 | (0.132) | 0.023 | (0.151) | 0.005 | (0.073) | 0.007 | (0.083) | 0.032 | (0.177) | 0.044 | (0.205) |
| HH Head is a Migrant | 0.044 | (0.206) | 0 | 0.000 | 0.01 | (0.101) | 0 | 0.000 | 0.086 | (0.280) | 0 | 0.000 |
| One Child is a Migrant | 0.187 | (0.390) | 0.2 | (0.400) | 0.136 | (0.342) | 0.142 | (0.349) | 0.25 | (0.433) | 0.272 | (0.445) |
| Two Children are Migrants | 0.158 | (0.365) | 0.175 | (0.380) | 0.049 | (0.215) | 0.057 | (0.232) | 0.29 | (0.454) | 0.322 | (0.467) |
| Other HH Member is Migrant | 0.023 | (0.150) | 0.022 | (0.147) | 0.011 | (0.107) | 0.01 | (0.100) | 0.037 | (0.188) | 0.037 | (0.189) |
| Migrated in the Past | 0.148 | (0.356) | 0.181 | (0.385) | 0.134 | (0.341) | 0.166 | (0.372) | 0.166 | (0.372) | 0.2 | (0.400) |
| No Migrants in HH | 0.612 | (0.487) | 0.609 | (0.488) | 0.799 | (0.401) | 0.793 | (0.405) | 0.386 | (0.487) | 0.382 | (0.486) |
| One Migrant in HH | 0.206 | (0.405) | 0.202 | (0.401) | 0.145 | (0.352) | 0.145 | (0.352) | 0.28 | (0.449) | 0.272 | (0.445) |
| Two Migrants in HH | 0.126 | (0.332) | 0.133 | (0.339) | 0.044 | (0.205) | 0.049 | (0.217) | 0.226 | (0.418) | 0.236 | (0.425) |
| More than Two Migrants in HH | 0.055 | (0.229) | 0.056 | (0.231) | 0.012 | (0.108) | 0.013 | (0.112) | 0.108 | (0.310) | 0.11 | (0.314) |
| Remittances (1000 CNY) | 3.823 | (7.201) | 3.781 | (7.258) | | | | | 8.432 | (8.689) | 8.471 | (8.850) |
| Remittances per Capita (1000 CNY) | 3.365 | (5.220) | 3.357 | (5.246) | | | | | 6.217 | (6.725) | 6.282 | (6.799) |
| Household Income (1000 CNY) | 22.363 | (23.895) | 22.203 | (24.652) | 27.37 | (29.048) | 26.785 | (30.170) | 16.326 | (13.244) | 16.519 | (13.200) |
| Household per C. Income (1000 CNY) | 14.962 | (15.429) | 15.351 | (16.520) | 17.727 | (18.362) | 18.05 | (19.789) | 11.628 | (9.914) | 12.003 | (10.272) |
| Wage Employment | 0.225 | (0.418) | 0.247 | (0.431) | 0.281 | (0.449) | 0.295 | (0.456) | 0.158 | (0.364) | 0.188 | (0.391) |
| Self-employment | 0.072 | (0.258) | 0.094 | (0.291) | 0.096 | (0.294) | 0.12 | (0.325) | 0.043 | (0.203) | 0.061 | (0.239) |

Source: RUMiC 2008.

## Table A3: Determinants of Subjective Well-Being in Rural China

| Dependent variable: GHQ-12 | Whole sample | | HH with remittances | | HH without remittances | |
|---|---|---|---|---|---|---|
| | All | HH head | All | HH head | All | HH head |
| *Economic Characteristics* | | | | | | |
| Log HH Income per Capita | 0.210** | 0.256*** | 0.206 | 0.293** | 0.237** | 0.243** |
| | (0.084) | (0.091) | (0.135) | (0.148) | (0.112) | (0.119) |
| Log Remittances per Capita | 0.175** | 0.176** | 0.198* | 0.293** | | |
| | (0.068) | (0.076) | (0.111) | (0.123) | | |
| Wageworker (D) | 1.111*** | 1.011*** | 0.086 | 0.523 | 1.568*** | 1.147*** |
| | (0.270) | (0.356) | (0.480) | (0.615) | (0.323) | (0.426) |
| Self-employed (D) | 1.129*** | 1.232*** | 0.241 | 0.739 | 1.502*** | 1.357*** |
| | (0.312) | (0.382) | (0.566) | (0.661) | (0.371) | (0.456) |
| Farmer (D) | 0.685*** | 0.618** | 0.783*** | 1.173*** | 0.553*** | 0.186 |
| | (0.165) | (0.269) | (0.265) | (0.443) | (0.205) | (0.327) |
| Hours of Work | -0.008* | -0.007 | 0.006 | 0.009 | -0.014*** | -0.015*** |
| | (0.004) | (0.005) | (0.008) | (0.009) | (0.005) | (0.006) |
| Land Size (Mu) | -0.022* | -0.023 | -0.038 | -0.053** | -0.017 | -0.011 |
| | (0.014) | (0.015) | (0.023) | (0.025) | (0.016) | (0.017) |
| House Size ($m^2$) | 0.507*** | 0.473*** | 0.36 | 0.516** | 0.530*** | 0.410** |
| | (0.140) | (0.151) | (0.227) | (0.253) | (0.177) | (0.190) |
| House Value (1000 CNY) | 0.078 | 0.125 | 0.111 | 0.104 | 0.102 | 0.164 |
| | (0.074) | (0.080) | (0.113) | (0.124) | (0.097) | (0.104) |
| *Socio-demographic Characteristics* | | | | | | |
| Age | -0.118*** | -0.009 | -0.088 | -0.002 | -0.131** | -0.008 |
| | (0.043) | (0.062) | (0.069) | (0.106) | (0.054) | (0.079) |
| Age squared | 0.001** | 0 | 0.001 | 0 | 0.001** | 0 |
| | 0.000 | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Male | 0.756*** | 0.901*** | 0.809*** | 1.144** | 0.734*** | 0.67 |
| | (0.191) | (0.325) | (0.286) | (0.444) | (0.258) | (0.479) |
| Married | 1.173*** | 1.538*** | 1.233*** | 1.579*** | 1.148*** | 1.529*** |
| | (0.255) | (0.315) | (0.389) | (0.474) | (0.331) | (0.418) |
| Good Health | -2.228*** | -2.012*** | -2.262*** | -2.317*** | -2.165*** | -1.797*** |
| | (0.121) | (0.140) | (0.189) | (0.221) | (0.158) | (0.181) |
| Average Health | -3.728*** | -3.349*** | -3.602*** | -3.368*** | -3.803*** | -3.389*** |
| | (0.154) | (0.179) | (0.227) | (0.263) | (0.211) | (0.246) |
| Poor Health | -8.143*** | -7.216*** | -8.052*** | -7.374*** | -8.174*** | -7.043*** |
| | (0.316) | (0.440) | (0.471) | (0.698) | (0.422) | (0.553) |
| Years of Education | 0.166*** | 0.140*** | 0.185*** | 0.131*** | 0.142*** | 0.139*** |
| | (0.018) | (0.024) | (0.028) | (0.039) | (0.024) | (0.031) |
| Height (cm) | 0.014 | -0.002 | 0.02 | 0.005 | 0.008 | -0.006 |
| | (0.010) | (0.013) | (0.015) | (0.020) | (0.013) | (0.017) |
| Weight (kg) | 0.014** | 0.019** | 0.007 | 0.008 | 0.023*** | 0.029** |
| | (0.007) | (0.009) | (0.010) | (0.013) | (0.009) | (0.011) |
| Urban Hukou | 0.113 | 0.408 | 0.841** | 1.845*** | -0.076 | 0.001 |
| | (0.217) | (0.250) | (0.395) | (0.484) | (0.256) | (0.287) |
| One Child | 0.086 | 0.665* | 0.515 | 1.036 | -0.296 | 0.358 |
| | (0.310) | (0.388) | (0.533) | (0.648) | (0.367) | (0.470) |
| Two Children | 0.181 | 0.564 | 0.396 | 0.686 | 0.044 | 0.518 |
| | (0.315) | (0.382) | (0.542) | (0.629) | (0.370) | (0.466) |
| More than Two Children | 0.079 | 0.559 | 0.515 | 0.949 | -0.297 | 0.268 |
| | (0.329) | (0.398) | (0.558) | (0.652) | (0.391) | (0.488) |
| # HH Members | 0.232*** | 0.226*** | 0.307*** | 0.386*** | 0.198*** | 0.147** |
| | (0.050) | (0.055) | (0.081) | (0.089) | (0.068) | (0.074) |
| *Migration Characteristics* | | | | | | |
| HH Head Migrant | 0.086 | | 0.317 | | -0.234 | |
| | (0.261) | | (0.300) | | (0.715) | |
| Spouse Migrant | -0.166 | -0.277 | -0.033 | -0.182 | 0.237 | -0.119 |
| | (0.414) | (0.394) | (0.465) | (0.447) | (0.905) | (0.879) |
| Has Migrated before 2007 | -0.076 | -0.228 | -0.056 | -0.3 | -0.063 | -0.186 |
| | (0.136) | (0.159) | (0.202) | (0.250) | (0.187) | (0.207) |
| One Migrant in HH | -0.644*** | -0.711*** | -0.928*** | -0.887*** | -0.561*** | -0.660*** |
| | (0.144) | (0.158) | (0.218) | (0.238) | (0.195) | (0.219) |
| Two Migrants in HH | -0.550*** | -0.566*** | -0.977*** | -0.986*** | -0.008 | 0.035 |
| | (0.195) | (0.213) | (0.259) | (0.281) | (0.329) | (0.362) |
| Three or More Migrants in HH | -1.136*** | -1.174*** | -1.665*** | -1.810*** | -1.369** | -1.462** |
| | (0.291) | (0.331) | (0.372) | (0.420) | (0.612) | (0.700) |
| Constant | 24.377*** | 22.707*** | 22.072*** | 20.442*** | 25.524*** | 23.798*** |
| | (1.872) | (2.564) | (3.005) | (4.172) | (2.401) | (3.284) |
| R-squared | 0.234 | 0.191 | 0.217 | 0.181 | 0.26 | 0.214 |
| Observations | 11624 | 6063 | 5270 | 2706 | 6354 | 3357 |

Notes: Models estimated using OLS. */**/*** indicate significance at the 10%, 5%, and 1% levels. Robust standard errors clustered at the household level are shown in parentheses. (D) indicates dummy variables. The model includes indicators for the 10 provinces as well as dummies indicating the relation to the household head (estimates are omitted). Per capita income and remittances are calculated using the modified OECD equivalence scale. GHQ-12 index obtained by summing the 12 answers of the General Health Questionnaire (GHQ).